Progress in Probability
70

Christoph Bandt
Kenneth Falconer
Martina Zähle
Editors

Fractal Geometry and Stochastics V

Birkhäuser

Birkhäuser

Progress in Probability
Volume 70

More information about this series at
http://www.springer.com/series/4839

Christoph Bandt • Kenneth Falconer •
Martina Zähle
Editors

Fractal Geometry and Stochastics V

Birkhäuser

Editors
Christoph Bandt
Institut für Mathematik und Informatik
Universität Greifswald
Greifswald, Germany

Kenneth Falconer
Mathematical Institute
University of St Andrews
St Andrews, United Kingdom

Martina Zähle
Institut für Mathematik
Friedrich-Schiller-Universität Jena
Jena, Germany

ISSN 1050-6977 ISSN 2297-0428 (electronic)
Progress in Probability
ISBN 978-3-319-18659-7 ISBN 978-3-319-18660-3 (eBook)
DOI 10.1007/978-3-319-18660-3

Library of Congress Control Number: 2015944516

Mathematics Subject Classification (2010): Primary: 28A80, 28A78, 28A75; Secondary: 37A, 37C, 60J, 60F, 60D, 37-06, 60-06

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.birkhauser-science.com)

Preface

The first conference of the series "Fractal Geometry and Stochastics," which took place in 1994, was the first meeting in Europe devoted to the mathematics of fractals. Since then, fractal structures and techniques have become well-established in many fields of mathematics, and conferences in the area have been organized in many countries. Held every 4 or 5 years, "Fractal Geometry and Stochastics" has continued to be a leading meeting in the field, gathering together the world's experts to discuss current developments, new trends, and open problems. For each of these conferences, the main contributions have been published by Birkhaeuser in their series *Progress in Probability*.

"Fractal Geometry and Stochastics V," with 123 participants from 23 countries, took place in Tabarz, Thuringia, Germany, from March 24 to 29, 2014. As in the previous meetings, we followed the principle of inviting representatives of very active areas of research, including new, young, and promising researchers. The main speakers were asked to write contributions for this volume. Most of them are nice introductions to the subjects, in the form of surveys with selected proofs. Some papers contain interesting original results. The volume is aimed both at newcomers in the field and to specialists.

We express our gratitude to the Deutsche Forschungsgemeinschaft for their financial support without which our conferences could not have been organized. We also thank a number of referees for their help in preparing this volume.

Christoph Bandt
Kenneth Falconer
Martina Zähle

Contents

Introduction

As a mathematical discipline, fractals have undergone a remarkable metamorphosis. They arose as curious counterexamples in topology around 1900. When Hausdorff defined fractal dimension in 1918, its significance was not understood by his contemporaries. It took several decades to develop geometric measure theory, both in its own right and as a tool for studying the geometry of highly irregular sets. In the 1970s Mandelbrot coined the word “fractal” and highlighted the potential of fractal structures for modeling nature. The advent of graphical computer interfaces boosted the field by attracting the attention of scientists, social scientists, and the general public but also by raising many deep mathematical questions. Thus, fractal techniques continue to grow in importance in both pure and applied fields.

The contributions to this volume reflect different aspects of this development in recent years. The authors are amongst the world’s leading experts in their fields, and they present their topics in an attractive and comprehensible manner. The book is divided into five parts, with papers ordered alphabetically within each part.

We begin with “Geometric Measure Theory”, a fundamental area. K. Falconer, J. Fraser, and X. Jin review the diverse and continuing research which has grown from Marstrand’s classical projection theorems from the 1950s. This survey is complemented by the paper of P. Shmerkin which shows how new techniques yield stronger projection results for classes of sets and measures with particular dynamical or arithmetic structure. The closely related concepts of scenery flow, and tangent distributions are used in A. Käenmäki’s paper to study questions of rectifiability and conical densities of general sets. The use of Minkowski functionals for describing anisotropy of fractals, motivated by needs of materials science, is discussed by P. Schönhöfer and K. Mecke.

“Self-Similar Fractals and Recurrent Structures” contains papers on particular recursive and self-similar constructions. K. Barański provides an accessible account of the recent proof of many cases of the long-standing conjecture for the Hausdorff dimension of Weierstrass functions and their generalizations. D. Feng and Y. Wang prove a new result characterizing tilings of the plane using translates of four tiles. M. Kesseböhmer and S. Zhu review recent results on the quantization dimension of measures, that is, how well measures can be approximated by discrete measures, with particular reference to self-similar and self-affine measures. Apollonian circle packings are discussed in M. Pollicott's paper which covers both contemporary and historical aspects, with particular emphasis on the rate of convergence of the circle radii. On the dynamical side, M. Rams studies the entropies of the Mather sets, defined in terms of Lyapunov exponents, for noncommutative dynamical systems.

Fractal structures in analysis and algebra are discussed in "Analysis and Algebra on Fractals." R. Grigorchuk, V. Nekrashevych, and Z. Šunić explain self-similar groups, along with related automata and fractals, and investigate the spectra of corresponding Schreier graphs. Self-similar graphs approximating certain fractals are also studied by P. Grabner, who calculates associated spectra of Laplacians by renormalization based on complex rational functions. This leads to fractal zeta functions, which are also central in the survey of M.L. Lapidus, G. Radunović, and D. Žubrinić, which introduces a new "distance" zeta function that extends the theory of complex dimensions of fractals to the multidimensional case.

Multifractals are the theme of "Multifractal Theory." Motivated by requirements of signal and image processing, S. Jaffard, P. Abry, C. Melot, R. Leonarduzzi, and H. Wendt introduce a new type of multifractal analysis, based on the concepts of a local p-exponent and lacunarity exponent of functions. J. Barral reviews recent work on the construction of measures and functions with prescribed multifractal characteristics.

"Random Constructions" conclude the volume. M. and E. Järvenpää review refined techniques to find the dimension of certain limsup sets defined by random subsets of d-dimensional tori. Then A. Telcs and M.E.-Nagy consider isoperimetric problems for random walks on weighted graphs.

Christoph Bandt
Kenneth Falconer
Martina Zähle

Part I
Geometric Measure Theory

Sixty Years of Fractal Projections

Kenneth Falconer, Jonathan Fraser, and Xiong Jin

Abstract Sixty years ago, John Marstrand published a paper which, among other things, relates the Hausdorff dimension of a plane set to the dimensions of its orthogonal projections onto lines. For a long time the paper attracted little attention. But over the past 30 years, Marstrand's projection theorems have become the prototype for many results in fractal geometry with numerous variants and applications and they continue to motivate leading research.

Keywords Fractals • Projections • Dimensions • Measure

Mathematics Subject Classification (2000). Primary 28A80; Secondary 28A78

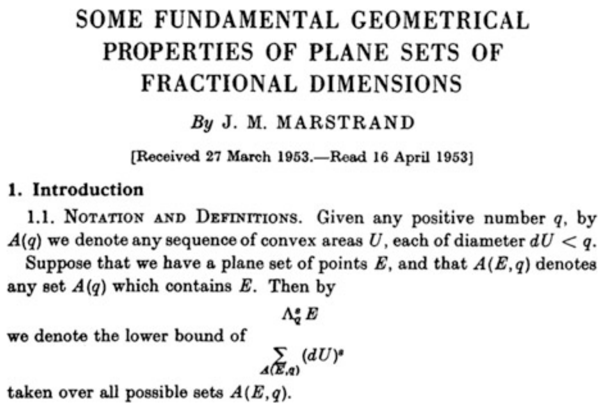
1 Marstrand's 1954 Paper

In 1954, John Marstrand's paper [56] 'Some fundamental geometrical properties of plane sets of fractional dimensions' was published in the Proceedings of the London Mathematical Society, see Fig. 1. The paper was essentially the work for his doctoral thesis at Oxford, which was heavily influenced by Abram Besicovitch, a Russian born mathematician and pioneer of geometric measure theory. For 25 years after its publication the paper attracted only limited attention, since then it has become one of the most frequently cited papers in the area now referred to as *fractal geometry*. Indeed, the paper was the first to consider the geometric properties of fractal dimensions.

K. Falconer (✉)
Mathematical Institute, University of St Andrews, North Haugh, St Andrews, Fife KY16 9SS, UK
e-mail: kjf@st-and.ac.uk

J. Fraser • X. Jin
School of Mathematics, The University of Manchester, Manchester M13 9PL, UK
e-mail: jonathan.fraser@manchester.ac.uk; xiong.jin@manchester.ac.uk

C. Bandt et al. (eds.), *Fractal Geometry and Stochastics V*, Progress in Probability 70, DOI 10.1007/978-3-319-18660-3_1


SOME FUNDAMENTAL GEOMETRICAL PROPERTIES OF PLANE SETS OF FRACTIONAL DIMENSIONS

By J. M. MARSTRAND

[Received 27 March 1953.—Read 16 April 1953]

1. Introduction

1.1. NOTATION AND DEFINITIONS. Given any positive number q, by $A(q)$ we denote any sequence of convex areas U, each of diameter $dU < q$.

Suppose that we have a plane set of points E, and that $A(E,q)$ denotes any set $A(q)$ which contains E. Then by

$$\Lambda_q^s E$$

we denote the lower bound of

$$\sum_{A(E,q)} (dU)^s$$

taken over all possible sets $A(E,q)$.

Fig. 1 Proceedings of the London Mathematical Society(3),**4** (1954), 257–302

The best-known results from the paper are the two Projection Theorems, stated below in Marstrand's wording, which relate the dimensions of sets in the plane to those of their orthogonal projections onto lines. Note that 'dimension' refers to Hausdorff dimension, and an 's-set' is a set that is measurable and of positive finite measure with respect to s-dimensional Hausdorff measure $\mathcal{H}^s$. 'Almost all directions' means all lines making angle θ with the x-axis except for a set of $\theta \in [0, \pi)$ of Lebesgue measure 0.

Theorem 1.1 *Any s-set whose dimension is greater than unity projects into a set of positive Lebesgue measure in almost all directions.*

Theorem 1.2 *Any s-set whose dimension does not exceed unity projects into a set of dimension s in almost all directions.*

The statements are followed by a remark that, by a result of Roy Davies [9], every Borel or analytic set of infinite s-dimensional Hausdorff measure contains an s-set. This allows the theorems to be expressed in terms of Hausdorff dimension, and this is the form in which they are now usually stated. We write $\dim_H$ for Hausdorff dimension, $\mathcal{L}$ for Lebesgue measure, and proj_θ for orthogonal projection of a set onto the line at angle θ to the x-axis, see Fig. 2.

Theorem 1.3 ([56]) *Let $E \subset \mathbb{R}^2$ be a Borel or analytic set. Then, for almost all $\theta \in [0, \pi)$,*

(i) $\dim_H \text{proj}_\theta E = \min\{\dim_H E, 1\}$,
(ii) $\mathcal{L}(\text{proj}_\theta E) > 0$ *if* $\dim_H E > 1$.

Since projection does not increase distances between points it follows easily from the definition of Hausdorff measure and dimension that $\dim_H \text{proj}_\theta E \leq \min\{\dim_H E, 1\}$ for all θ, but the opposite inequality is much more intricate. Marstrand's proofs depend heavily on plane geometry and measure theory, with, for example, careful estimates of the measures of narrow strips in various directions. As John Marstrand once remarked, analysis essentially consists of integrating functions

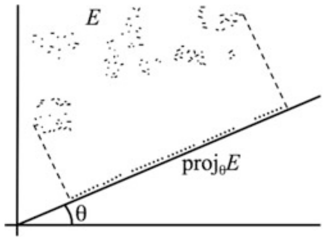


Fig. 2 Projection of a set E onto a line in direction θ

in different ways and applying Fubini's theorem – but it may be difficult to find an appropriate function. The proofs in this paper illustrate this well.

It is worth mentioning that Marstrand's paper [56] includes a nice, but often forgotten, extension to the theorems, that the same exceptional directions can apply to subsets of the given s-set that are of positive measure. In the following statement from the paper $|\ \ |$ denotes Lebesgue measure.

Proposition 1.4 *If E is an s-set and $s > 1$, then for almost all angles θ, all s-sets A which are contained in E satisfy* $|\mathrm{proj}_\theta A| > 0$.

Although Marstrand's paper is most often cited for the projection theorems, its 46 pages contain a great deal more, much of which anticipated other directions in fractal geometry.

- *Dimension of the intersection of sets with lines.* E.g. Almost every line through $\mathcal{H}^s$-almost every point of an s-set E ($s > 1$) intersects E in a set of dimension $s - 1$ and finite $s - 1$-dimensional measure.
- *Construction of examples with particular projection properties.* E.g. For $1 < s < 2$ there exists an s-set which projects onto a set of dimension $s - 1$ in continuum many directions in every sector.
- *Dimension of exceptional sets.* The dimension of the set of points from which an irregular 1-set (see Sect. 4) has projection of positive Lebesgue measure is at most 1.
- *Densities of s-sets.* The density $\lim_{r\to 0} \mathcal{H}^s(E \cap B(x,r))/(2r)^s$ of an s-set $E \subset \mathbb{R}^2$ can exist and equal 1 for $\mathcal{H}^s$-almost all x only if $s = 0, 1$ or 2. ($B(x,r)$ denotes the disc of centre x and radius r.)
- *Angular densities.* Bounds are given for densities defined in segments emanating from points of an s-set.
- *Weak tangents.* For $1 < s < 2$ an s-set fails to have a weak tangent (with an appropriate definition) almost everywhere.

This area of research is a central part of what is now termed fractal geometry. This paper will survey the vast range of mathematics related to projections of sets that has developed over the past 60 years and which might be regarded as having its genesis in Marstrand's 1954 paper.

2 The Potential-Theoretic Approach

By virtue of the fact that an orthogonal projection is a Lipschitz map, we invariably have $\dim_H \mathrm{proj} E \leq \min\{m, \dim_H E\}$ for every set $E \subset \mathbb{R}^n$ and projection $\mathrm{proj} : \mathbb{R}^n \to V$ onto every m-dimensional subspace V, a fact that should be borne in mind throughout this article. It is inequalities in the opposite direction that require more work to establish. (However, a particularly straightforward situation is that for a connected set $E \subset \mathbb{R}^2$, both $\dim_H E \geq 1$ and $\dim_H \mathrm{proj} E = 1$ for projections onto lines in all directions with at most one exception.) Throughout this article we will always assume that the sets E being projected are Borel or analytic – pathological constructions show that dimension ceases to have useful geometric properties if completely general sets are considered.

Marstrand's proofs of his projection theorems were geometrically complicated and not particularly conducive to extension or generalization. But in 1968 Kaufman [50] gave new proofs of Theorem 1.1(i) using potential theory and of Theorem 1.1(ii) using a Fourier transform method. This provided a rather more accessible approach, leading eventually to many generalizations and extensions. Kaufman's proofs depend on the following characterization of Hausdorff dimension in terms of an energy integral:

$$\dim_H E = \sup\Big\{s : E \text{ supports a positive finite measure } \mu \text{ such that } \int\int \frac{d\mu(x)d\mu(y)}{|x-y|^s} < \infty\Big\}. \tag{2.1}$$

Thus if $E \subset \mathbb{R}^2$ and $s < \dim_H E$ where $0 < s < 1$, we may find a measure μ supported by E such that $\int\int \frac{d\mu(x)d\mu(y)}{|x-y|^s} < \infty$. Write μ_θ for the projection of μ onto the line in direction θ, so $\int_{-\infty}^{\infty} f(t)d\mu_\theta(t) = \int_E f(x\cdot\theta)d\mu(x)$ for continuous f, where we identify θ with a unit vector in the direction θ. Then

$$\begin{aligned}\int_0^\pi \Big[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{d\mu_\theta(t)d\mu_\theta(u)}{|t-u|^s}\Big]d\theta &= \int_0^\pi \Big[\int_E\int_E \frac{d\mu(x)d\mu(y)}{|x\cdot\theta - y\cdot\theta|^s}\Big]d\theta \\ &= \int_E\int_E\int_0^\pi \frac{d\theta}{|u_{x-y}\cdot\theta|^s}\frac{d\mu(x)d\mu(y)}{|x-y|^s} \\ &\leq c\int_E\int_E \frac{d\mu(x)d\mu(y)}{|x-y|^s} < \infty\end{aligned} \tag{2.2}$$

where u_w denotes the unit vector $w/|w|$ and $\int_0^\pi |u_{x-y}\cdot\theta|^{-s}d\theta = c < \infty$.

Hence for almost all θ, $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{d\mu_\theta(t)d\mu_\theta(u)}{|t-u|^s} < \infty$, so, since μ_θ is supported by $\mathrm{proj}_\theta E$, we conclude from the characterization (2.1) that $\dim_H \mathrm{proj}_\theta E \geq s$. This is true for all $s < \dim_H E$, so $\dim_H \mathrm{proj}_\theta E \geq \dim_H E$ for almost all θ.

For the case where $1 < s < 2$, a variant of this argument shows that

$$\int_0^{\pi} \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\widehat{\mu_\theta}(u)|^2 \right] du < \infty$$

where $\widehat{\mu_\theta}$ is the Fourier transform of μ_θ from which it follows that μ_θ is absolutely continuous with respect to Lebesgue measure with L^2 density, so in particular has support of positive Lebesgue measure.

In 1975 Mattila [57] used potential theoretic methods to obtain the natural extension of these theorems to projections from higher dimensional spaces to subspaces. For $1 \le m < n$, and V an m-dimensional subspace of $\mathbb{R}^n$, let proj_V denote orthogonal projection onto V. These subspaces form the Grassmanian $G(n,m)$, an $m(n-m)$-dimensional compact manifold which carries a natural invariant measure, locally equivalent to $m(n-m)$-dimensional Lebesgue measure.

Theorem 2.1 ([57]) *Let $E \subset \mathbb{R}^n$ be a Borel or analytic set. Then, for almost all $V \in G(n,m)$,*

(i) $\dim_{\rm H} \text{proj}_V E = \min\{\dim_{\rm H} E, m\}$.
(ii) $\mathcal{L}^m(\text{proj}_V E) > 0$ *if* $\dim_{\rm H} E > m$, *where $\mathcal{L}^m$ denotes m-dimensional Lebesgue measure on V.*

3 Exceptional Sets of Projections

We can deduce rather more from Kaufman's proof above. Let $E \subset \mathbb{R}^2$ and $0 < s < \dim_{\rm H} E < 1$. Let $T = \{\theta : \dim_{\rm H} \text{proj}_\theta E < s\}$. If $\dim_{\rm H} T > s$ then it can be shown that we may find a measure ν supported by T such that $\int_T |\mathbf{u} \cdot \theta|^{-s} d\nu(\theta) \le c < \infty$ for every unit vector $\mathbf{u}$. If we integrate with respect to ν instead of Lebesgue measure in (2.2) we still get a finite triple integral, and so for ν-almost all $\theta \in T$ the s-energy of μ_θ is finite and $\dim_{\rm H} \text{proj}_\theta E \ge s$, a contradiction. It follows that if $E \subset \mathbb{R}^2$ and $0 \le s < \dim_{\rm H} E < 1$ then

$$\dim_{\rm H}\{\theta : \dim_{\rm H} \text{proj}_\theta E < s\} \le s.$$

Thus the set of θ for which the projections have much smaller dimension than that of the set is correspondingly small. Indeed, the dimension of a projection is rarely less than half that of the set. As Bourgain [8] and Oberlin [63] showed, again when $E \subset \mathbb{R}^2$ and $\dim_{\rm H} E < 1$,

$$\dim_{\rm H}\{\theta : \dim_{\rm H} \text{proj}_\theta E < \tfrac{1}{2} \dim_{\rm H} E\} = 0.$$

For $E \subset \mathbb{R}^2$ and $\dim_H E > 1$, the greater the 'excess dimension' $\dim_H E - 1$ the smaller the set of θ where Marstrand's conclusion fails. To be more precise:

$$\dim_H\{\theta : \mathcal{L}(\text{proj}_\theta E) = 0\} \leq 2 - \dim_H E.$$

This was first proved in [14] and all known proofs depend on Fourier transforms.

Not surprisingly there are higher dimensional analogues of these bounds on the dimensions of the exceptional set, that is the set of $V \in G(n, m)$ for which the conclusions of Theorem 2.1 fail. These are summerised in the following inequalities, written for comparison with $m(n-m)$, the dimension of the Grassmanian $G(n, m)$, see [51, 57–59] for more details.

Theorem 3.1 *Let $E \subset \mathbb{R}^n$ be a Borel or analytic set.*

(i) If $0 < s < \dim_H E \leq m$ *then*

$$\dim_H\{V \in G(n, m) : \dim_H \text{proj}_V E < s\} \leq m(n-m) - (m-s);$$

(ii) If $\dim_H E \geq m$ *then*

$$\dim_H\{V \in G(n, m) : \dim_H \text{proj}_V E < s\} \leq m(n-m) - (\dim_H E - s);$$

(iii) If $\dim_H E > m$ *then*

$$\dim_H\{V \in G(n, m) : \mathcal{L}^m(\text{proj}_V E) = 0\} \leq m(n-m) - (\dim_H E - m);$$

(iv) If $\dim_H E > 2m$ *then*

$$\dim_H\{V \in G(n, m) : \text{proj}_V E \textit{ has empty interior}\} \leq m(n-m) - (\dim_H E - 2m).$$

4 Sets of Integer Dimension

Marstrand was the first person to consider the effect of projection on the numerical value of the dimension, but his paper also includes a few results on projections of s-sets in the 'critical case' where $s = 1$. This case had been studied in great detail somewhat earlier by Besicovitch around the 1930s [5–7] who showed that 1-sets or 'linearly-measurable sets' in the plane could be decomposed into a regular part and an irregular part, using local densities $D(x) = \lim_{r \to 0} \mathcal{H}^1(E \cap B(x, r))/2r$.

The *regular* part consists of those x where the limit $D(x)$ exists with $D(x) = 1$, and the *irregular* part is formed by the remaining points. Besicovitch showed that, to within a set of measure 0, the regular part is 'curve-like', that is a subset of a countable collection of rectifiable curves. On the other hand, the irregular part is 'dust-like' intersecting every rectifiable curve in length 0. Using intricate geometrical arguments, Besicovitch obtained the following projection theorem.

Theorem 4.1 (**[7]**) *Let $E \subset \mathbb{R}^2$ be a* 1*-set.*

(i) If E is regular then $\mathcal{L}(\text{proj}_\theta E) > 0$ for all $\theta \in [0, \pi)$ except perhaps for a single value of θ.
(ii) If E is irregular then $\mathcal{L}(\text{proj}_\theta E) = 0$ for almost all $\theta \in [0, \pi)$.

The natural higher dimensional versions of Theorem 4.1, with appropriate definitions of regular and irregular sets, were obtained by Federer [27, 28].

If E is measurable and of σ-finite $\mathcal{H}^1$ measure, it follows from Theorem 4.1 that $\mathcal{L}(\text{proj}_\theta E)$ is either 0 for almost all θ or positive for almost all θ, by decomposing E into countably many 1-sets. However if $\dim_{\text{H}} E = 1$ but E is not σ-finite then strange things can occur: we can find a set E whose projections are, to within Lebesgue measure 0, anything we like.

Theorem 4.2 (**[9, 15]**) *For each $\theta \in [0, \pi)$ let E_θ be a given subset of the line through the origin of $\mathbb{R}^2$ in direction θ, such that $\bigcup_{0 \le \theta < \pi} E_\theta$ is plane Lebesgue measurable. Then there exists a Borel set $E \subset \mathbb{R}^2$ such that, for almost all θ, $\mathcal{L}(E_\theta \bigtriangleup \text{proj}_\theta E) = 0$ where $\bigtriangleup$ denotes symmetric difference, in other words $\text{proj}_\theta E$ differs from the prescribed set E_θ by a set of negligible length.*

Theorem 4.2 may be obtained by dualising a result of Davies [9, 10] on covering a plane set by lines without increasing its plane Lebesgue measure. It was proved directly, along with the natural higher dimension analogues, in [15]. For projections from $\mathbb{R}^3$ to $\mathbb{R}^2$ this has become known as the 'digital sundial theorem': Given a subset E_V of each 2-dimensional subspace V of $\mathbb{R}^3$ (with a measurability condition), there exists a Borel set $E \subset \mathbb{R}^3$ such that, for almost all subspaces V, $\mathcal{L}^2(E_V \bigtriangleup \text{proj}_V E) = 0$. Thus, in theory at least, there is a set in space such that the shadow cast by the sun gives the thickened digits of the time at any instant, see Fig. 3.

5 Packing Dimensions

Packing measures and packing dimension were introduced by Tricot [84] in 1982 as a sort of dual to their Hausdorff counterparts, see [17, 58]. Whilst packing measures require an extra step in their definition, the gap of over 60 years between the two concepts seems very surprising with hindsight. Nowadays, however, every problem that involves Hausdorff dimension is almost routinely studied in terms of packing dimension as well. Projection theorems are no exception, but the dimensional relationships turn out to be more complicated in the packing dimension case.

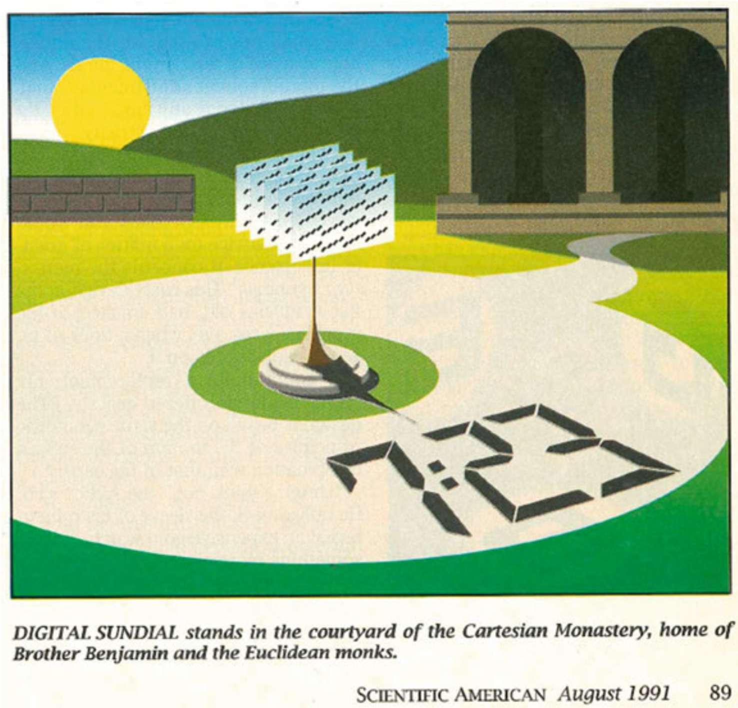


Fig. 3 A digital sundial – drawing by Andrew Christie

Järvenpää [43] constructed compact sets $E \subset \mathbb{R}^n$ with $\dim_{\mathrm{P}} E$ taking any prescribed value in $(0, n]$ such that $\dim_{\mathrm{P}} \mathrm{proj}_V E = \dim_{\mathrm{P}} E \big/ \big(1+(1/m-1/n) \dim_{\mathrm{P}} E\big)$ for all $V \in G(n,m)$. This is essentially the least value that can be obtained, that is

$$\frac{\dim_{\mathrm{P}} E}{1 + (1/m - 1/n) \dim_{\mathrm{P}} E} \le \dim_{\mathrm{P}} \mathrm{proj}_V E \le \min\{\dim_{\mathrm{P}} E, m\}$$

for almost all $V \in G(n,m)$, see [19]. For packing dimensions of projections of measures, rather than sets, this lower bound was refined to incorporate both the Hausdorff and packing dimensions of the measure, see [23].

These inequalities raised the question of whether $\dim_{\mathrm{P}} \mathrm{proj}_V E$ takes a common value for almost all subspaces V and this was answered affirmatively with the introduction of 'dimension profiles' [20]. The packing dimension profile $\dim_{\mathrm{P}}^s E$ of a set $E \subset \mathbb{R}^n$ reflects how E appears when viewed in an s-dimensional setting. For $s > 0$ the *s-dimensional packing dimension profile* of a measure μ on $\mathbb{R}^n$ with bounded support is defined in terms of local densities of measures with respect to a kernel of the form $\min\{1, r^s/|x-y|^s\}$:

$$\dim_{\mathrm{P}}^s \mu = \sup\left\{ t \geq 0 : \liminf_{r \searrow 0} r^{-t} \int \min\left\{1, \frac{r^s}{|x-y|^s}\right\} d\mu(y) < \infty \right.$$
$$\left. \text{for } \mu\text{-almost all } x \in \mathbb{R}^n \right\}.$$

This leads to the *s-dimensional packing dimension profile* of a set $E \subset \mathbb{R}^n$

$$\dim_{\mathrm{P}}^s E = \sup\left\{ \dim_{\mathrm{P}}^s \mu : \mu \text{ is a finite compactly supported measure on } E \right\},$$

see [20]. The profiles generalize packing dimensions, since $\dim_{\mathrm{P}}^n E = \dim_{\mathrm{P}} E$ for $E \subset \mathbb{R}^n$. The profiles may also be expressed in terms of measures defined by weighted coverings, see [38, 53].

Theorem 5.1 ([20]) *Let $E \subset \mathbb{R}^n$ be a Borel or analytic set. Then, for almost all $V \in G(n, m)$,*

$$\dim_{\mathrm{P}} \mathrm{proj}_V E = \dim_{\mathrm{P}}^m E.$$

There is a certain parallel with Hausdorff dimensions, where one might define a dimension profile simply as $\dim_{\mathrm{H}}^s E = \min\{s, \dim_{\mathrm{H}} E\}$ which, by Marstrand's theorem, gives the Hausdorff dimension of projections onto almost all s-dimensional subspaces.

As well as giving the almost sure packing dimension of the projections, the profiles provide upper bounds for the dimension of the exceptional set of directions for which the packing dimension falls below the almost sure value see [20] and also [68].

Since their introduction, packing dimension profiles have cropped up in other contexts, notably to give the almost sure packing dimension of images of sets under fractional Brownian motion [53, 85].

6 Projections in Restricted Directions

A general question that has been around for many years is under what circumstances we can get projection theorems for projections onto families of lines or subspaces that form proper subsets of $V(n, m)$. For instance, if $\{\theta(t) : t \in P\}$ is a smooth curve or submanifold of $V(n, m)$ smoothly parameterized by a set $P \subset \mathbb{R}^k$, then what can we conclude about $\dim_{\mathrm{H}} \mathrm{proj}_{\theta(t)} E$ for $\mathcal{L}^k$-almost all $t \in P$, where $\mathcal{L}^k$ is k-dimensional Lebesgue measure?

For a simple example, it follows easily from Theorem 3.1 (ii)–(iii) that if $\{\theta(t) : 0 \leq t \leq 1\}$ is a smoothly parameterized curve of directions in $\mathbb{R}^3$ (i.e. a curve in $V(3, 1)$), then for almost all $0 \leq t \leq 1$ we have $\dim_{\mathrm{H}} \mathrm{proj}_\theta E \geq \min\{\dim_{\mathrm{H}} E - 1, 1\}$

and if $\dim_{\rm H} E > 2$ then $\mathcal{L}^1(\text{proj}_{\theta(t)}E) > 0$, where $\text{proj}_{\theta(t)}$ denotes projection onto the line in direction $\theta(t)$.

The following lower bounds were obtained Järvenpää, Järvenpää and Keleti [44] for parameterized families of projections from $\mathbb{R}^n$ to m-dimensional subspaces, see also [45]. For $0 < k < m(n-m)$ define the integers

$$p(l) = n - m - \left\lfloor \frac{k - l(n-m)}{m-l} \right\rfloor \qquad (l = 0, 1, \ldots, m-1),$$

where the 'floor' symbol '$\lfloor x \rfloor$' denotes the largest integer no greater than x.

Theorem 6.1 ([44]) *Let $P \subset \mathbb{R}^k$ be an open parameter set and let $E \subset \mathbb{R}^n$ be a Borel or analytic set. Let $\{V(t) \subset G(n,m) : t \in P\}$ be a family of subspaces such that V is C^1 with the derivative $D_t V(t)$ injective for all $t \in P$. Then, for all $l = 0, 1, \ldots, m$ and $\mathcal{L}^k$-almost all $t \in P$,*

$$\dim_{\rm H} \text{proj}_{V(t)}E \geq \begin{cases} \dim_{\rm H} E - p(l) & \text{if } p(l) + l \leq \dim_{\rm H} E \leq p(l) + l + 1 \\ l + 1 & \text{if } p(l) + l + 1 \leq \dim_{\rm H} E \leq p(l+1) + l + 1. \end{cases}$$

Moreover, if $\dim_{\rm H} E > p(m-1) + m$ then $\mathcal{L}^m(\text{proj}_{V(t)}E) > 0$ for $\mathcal{L}^k$-almost all $t \in P$.

These are the best possible bounds for general parameterized families of projections. The same paper [44] includes generalizations of these results to smoothly parameterized families of C^2-mappings.

Better lower bounds may be obtained if there is curvature in the mapping $s \mapsto V(s)$. This is a difficult area, and work to date mainly concerns projections from $\mathbb{R}^3$ to lines and planes. Let $\theta : (0,1) \to S^2$ be a family of directions given by a C^3-function θ, where S^2 is the 2-sphere embedded in $\mathbb{R}^3$. We say that the family of directions is *non-degenerate* if

$$\text{span}\,\{\theta(t), \theta'(t), \theta''(t)\} = \mathbb{R}^3 \qquad \text{for all } t \in (0,1).$$

The following theorem was proved by recently by Fässler and Orponen [26].

Theorem 6.2 ([26]) *Let $E \subset \mathbb{R}^3$ be a Borel or analytic set, let $\theta(t)$ be a non-degenerate family of directions, and let $\text{proj}_{\theta(t)}$ denote projection onto the line in direction θ. Then, for almost all $t \in (0,1)$,*

$$\dim_{\rm H} \text{proj}_{\theta(t)}E \geq \min\{\dim_{\rm H} E, \tfrac{1}{2}\}. \tag{6.1}$$

It is conjectured that $\frac{1}{2}$ can be replaced by 1 in (6.1) and this is verified where E is a self-similar set without rotations in [26], a paper that also contains estimates for packing dimensions of projections.

The following bounds have been established for projections onto planes in $\mathbb{R}^3$ in the non-degenerate case. The conjectured lower bound is $\min\{\dim_{\rm H} E, 2\}$ and the bound $\min\{\dim_H E, 1\}$ for all values of $\dim_H E$ was obtained in [26]. The further improvements stated below come from Fourier restriction methods [64, 65].

Theorem 6.3 ([26, 64, 65]) *Let $E \subset \mathbb{R}^3$ be a Borel or analytic set, let $\theta(t)$ be a non-degenerate family of directions, and let $\text{proj}_{V_\theta(t)}$ denote projection onto the plane perpendicular to direction θ. Then, for almost all $t \in (0, 1)$,*

$$\dim_H \text{proj}_{V_\theta(t)} E \geq \begin{cases} \min\{\dim_H E, 1\} & \text{if } 0 \leq \dim_H E \leq \frac{4}{3} \\ \frac{3}{4} \dim_H E & \text{if } \frac{4}{3} \leq \dim_H E \leq 2 \\ \min\{\dim_H E - \frac{1}{2}, 2\} & \text{if } 2 \leq \dim_H E \leq 3 \end{cases}.$$

Orponen [71] also showed that there exist numbers $\sigma(\lambda) > 1$ defined for $\lambda > 1$, and increasing with λ, such that if $\dim_H E > 1$ then $\dim_H \text{proj}_{V_\theta(t)} E \geq \sigma(\dim_H E)$ for almost all t.

Estimates for packing dimensions of projections may be found in [26]. The introduction of the paper [71] provides a recent overview of this area.

7 Generalized Projections

The projection theorems are a special case of much more general results. The essential property in Kaufman's proof in Sect. 2 is that the integral over the parameter θ satisfies $\int |\text{proj}_\theta x - \text{proj}_\theta y|^{-s} d\theta \leq c|x - y|^{-s}$; such a condition can hold for many other parameterized families of mappings as well as for proj_θ.

Thus for $X \subset \mathbb{R}^n$ a compact domain, consider a family of maps $\pi_\theta : X \to \mathbb{R}^m$ for θ in an open parameter set $P \subset \mathbb{R}^k$. Assume that the derivatives with respect to θ, $D_\theta \pi_\theta(x)$ exist and are bounded.

Let

$$\Phi_\theta(x, y) = \frac{|\pi_\theta(x) - \pi_\theta(y)|}{|x - y|}.$$

The family $\{\pi_\theta : \theta \in P\}$ is *transversal* if there is a constant c such that

$$|\Phi_\theta(x, y)| \leq c \implies \det\left(D_\theta \Phi_\theta(x, y)(D_\theta \Phi_\theta(x, y))^T\right) \geq c$$

for $\theta \in P$ and $x, y \in X$, where D_θ denotes the derivative with respect to θ and T denotes the transpose of a matrix. (A form of transversality was first introduced in [75]). This condition implies that if $\theta \in P$ is such that $\Phi_\theta(x, y)$ is small, then $\Phi_\theta(x, y)$ must be varying reasonably fast as θ changes in a direction perpendicular to the kernel of the derivative matrix.

By generalizing beyond recognition earlier arguments involving potential theory and Fourier transforms, Peres and Schlag [72] obtained theorems such as the following for a transversal family of generalized projections; compare Theorem 3.1.

Theorem 7.1 ([72]) *For $X \subset \mathbb{R}^n$ and $P \subset \mathbb{R}^k$, let $\{\pi_\theta : X \to \mathbb{R}^m : \theta \in P\}$ be a transversal family and let $E \subset X$ be a Borel set.*

(i) If $0 < t < \dim_H E \leq m$ *then*

$$\dim_H\{\theta \in P : \dim_H \pi_\theta E < t\} \leq k - (m - t);$$

(ii) If $\dim_H E > m$ *then*

$$\dim_H\{\theta \in P : \dim_H \pi_\theta E < t\} \leq k - (\dim_H E - t);$$

(iii) If $\dim_H E > m$ *then*

$$\dim_H\{\theta \in P : \mathcal{L}^m(\pi_\theta E) = 0\} \leq k - (\dim_H E - m);$$

(iv) If $\dim_H E > 2m$ *then*

$$\dim_H\{\theta \in P : \pi_\theta E \textit{ has empty interior}\} \leq n - \dim_H E + 2.$$

This powerful result has been applied to many situations, including Bernoulli convolutions, sums of Cantor sets and pinned distance sets, see [72]. For a recent treatment of transversality, see [62].

Leikas [55] has used transversality to extend the packing dimension conclusions of Sect. 5 to families of mappings between Riemannian manifolds where the dimension profiles again play a central role.

8 Projections of Self-Similar and Self-Affine Sets

One of the difficulties with the projection theorems is that they tell us nothing about the dimension or measure of the projection in any given direction. There has been considerable recent interest in examining the dimensions of projections in specific directions for particular sets or classes of sets, and especially in finding sets for which the conclusions of Marstrand's theorems are valid for all, or virtually all, directions. Of particular interest are self-similar sets which we consider briefly here; there is a very nice and much more detailed account of the area by Shmerkin [80] elsewhere in this volume.

Recall that an *iterated function system* (IFS) is a family of contractions $\{f_1, \ldots, f_k\}$ with $f_i : \mathbb{R}^n \to \mathbb{R}^n$. An IFS determines a unique non-empty compact $E \subset \mathbb{R}^d$ such that

$$E = \bigcup_{i=1}^{k} f_i(E), \tag{8.1}$$

called the *attractor* of the IFS, see [17, 42]. If the f_i are all similarities, that is of the form

$$f_i(x) = r_i O_i(x) + a_i, \tag{8.2}$$

where $0 < r_i < 1$ is the similarity ratio, O_i is an orthonormal map, i.e. a rotation or reflection, and a_i is a translation, then E is termed *self-similar*. An IFS of similarities satisfies the *strong separation condition* (SSC) if the union (8.1) is disjoint, and the *open set condition* (OSC) if there is a non-empty open set U with $\cup_{i=1}^k f_i(U) \subset U$ with this union disjoint. If either SSC or OSC hold then $\dim_H E = s$, where s is the *similarity dimension* given by $\sum_{i=1}^k r_i^s = 1$, where r_i is the similarity ratio of f_i, and moreover $0 < \mathcal{H}^s(E) < \infty$. The *rotation group* $G = \langle O_1, \ldots, O_k \rangle$ generated by the orthonormal components of the similarities plays a crucial role in the behaviour of the projections of self-similar sets.

It is easy to construct self-similar sets with a finite rotation group G for which the conclusions of Marstrand's theorem fail in certain directions. For example, let $f_1, \ldots, f_4$ be homotheties (that is similarities with O_i the identity in (8.2)) of ratio $0 < r < \frac{1}{4}$ that map the unit square S into itself, each f_i fixing one of the four corners. Then $\dim_H E = -\log 4/\log r$, but the projections of E onto the sides of the square have dimension $-\log 2/\log r$ and onto the diagonals of S have dimension $-\log 3/\log r$, a consequence of the alignment of the component squares $f_i(S)$ under projection. There is a similar reduction in the dimension of projection in direction θ whenever $\mathrm{proj}_\theta(f_{i_1} \circ \cdots \circ f_{i_k}(S)) = \mathrm{proj}_\theta(f_{j_1} \circ \cdots \circ f_{j_k}(S))$ for distinct words $i_1, \ldots, i_k$ and $j_1, \ldots, j_k$.

Kenyon [52] conducted a detailed investigation of the projections of the 1-dimensional Sierpiński gasket $E \subset \mathbb{R}^2$, that is the self-similar set defined by the similarities

$$f_1(x,y) = (\tfrac{1}{3}x, \tfrac{1}{3}y),\ f_2(x,y) = (\tfrac{1}{3}x + \tfrac{2}{3}, \tfrac{1}{3}y),\ f_3(x,y) = (\tfrac{1}{3}x, \tfrac{1}{3}y + \tfrac{2}{3}).$$

He showed that the projection of E onto a line making an angle to the x-axis with tangent p/q with has dimension strictly less than 1 if $p + q \not\equiv 0 \pmod 3$, but if $p + q \equiv 0 \pmod 3$ then the projection has non-empty interior. For irrational directions he proved that the projections have Lebesgue measure 0 and Hochman [34] complemented this by showing that they nevertheless have Hausdorff dimension 1.

In fact, when the rotation group is finite, there are always some projections for which direct overlapping of the projection of components of the usual iterated construction leads to a reduction in dimension, as the following theorem of Farkas shows.

Theorem 8.1 ([25]) *If $E \subset \mathbb{R}^n$ is self-similar with finite rotation group G and similarity dimension s, then $\dim_H \mathrm{proj}_V E < s$ for some $V \in G(n,m)$. In particular if E satisfies OSC and $0 < \dim_H E < m$ then $\dim_H \mathrm{proj}_V E < \dim_H E$ for some V.*

A rather different situation occurs if the IFS has *dense rotations*, that is the rotation group G is dense in the full group of rotations $SO(n,\mathbb{R})$ or in the group of isometries $O(n,\mathbb{R})$. Note that an IFS of similarities of the plane has dense rotations if at least one of the rotations in the group is an irrational multiple of π.

Theorem 8.2 ([35, 74]) *If $E \subset \mathbb{R}^n$ is self-similar with dense rotations then*

$$\dim_{\mathrm{H}} \mathrm{proj}_V E = \min\{\dim_{\mathrm{H}} E, m\} \text{ for all } V \in G(n,m). \tag{8.3}$$

More generally, $\dim_{\mathrm{H}} g(E) = \min\{\dim_{\mathrm{H}} E, m\}$ *for all C^1 mappings $g : E \to \mathbb{R}^m$ without singular points, that is maps with non-singular derivative matrix.*

Peres and Shmerkin [74] proved (8.3) in the plane without requiring any separation condition on the IFS. To show this they set up a discrete version of Marstrand's projection theorem to construct a tree of intervals in the subspace (line) V followed by an application of Weyl's equidistribution theorem. Hochman and Shmerkin [35] proved the theorem in higher dimensions, including the extension to C^1 mappings, for E satisfying the open set condition. Their proof uses the CP-chains of Furstenberg [31, 32], see also [33], and has three main ingredients: the lower semicontinuity of the expected Hausdorff dimension of the projection of a measure with respect to its 'micromeasures', Marstrand's projection theorem, and the invariance of the dimension of projections under the action of the rotation group.

That the open set condition is not essential follows since, for all $\epsilon > 0$, we can use a Vitali argument to set up a new IFS, consisting of compositions of the f_i, that satisfies SSC, with attractor $E' \subset E$ such that $\dim_{\mathrm{H}} E' > \dim_{\mathrm{H}} E - \epsilon$; we can also ensure that the new IFS has dense rotations if the original one has, see [21, 25, 69, 74].

It is also natural to ask about the Lebesgue measures of the projections of self-similar sets. We have seen examples of self-similar sets E of Hausdorff dimension $s < m$ with finite rotation group and satisfying SSC such that $\mathcal{H}^s(\mathrm{proj}_V E)$ is positive for some subspaces V and 0 for others. For dense rotations, the situation is clear cut: the following theorem was proved by Eroğlu [13] in the plane case for projections when OSC is satisfied, and for more general mappings with the separation condition removed by Farkas [25].

Theorem 8.3 ([25]) *Let $E \subset \mathbb{R}^n$ be the self-similar attractor of an IFS with dense rotations, with* $\dim_{\mathrm{H}} E = s$*. Then* $\mathcal{H}^s(\mathrm{proj}_V E) = 0$ *for all $V \in G(n,m)$. More generally,* $\mathcal{H}^s(g(E)) = 0$ *for all C^1 mappings $g : E \to \mathbb{R}^n$ without singular points.*

In the dense rotation case, if $\dim_{\mathrm{H}} E > m$ then $\dim_{\mathrm{H}} \mathrm{proj}_V E = m$ for all $V \in G(n,m)$ by Theorem 8.2, but we might hope from the second part of Marstrand's theorem that the projections also have positive Lebesgue measure. Shmerkin and Solomyak showed that this is very nearly so in the plane.

Theorem 8.4 ([81]) *Let $E \subset \mathbb{R}^2$ be the self-similar attractor of an IFS with dense rotations with* $1 < \dim_{\mathrm{H}} E < 2$*. Then* $\mathcal{L}^1(\mathrm{proj}_\theta E) > 0$ *for all θ except for a set of θ of Hausdorff dimension 0.*

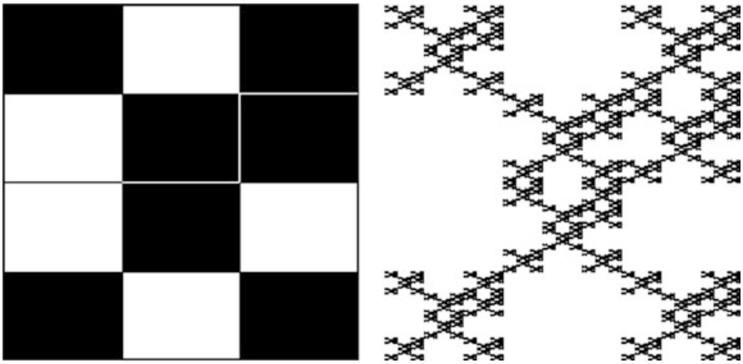


Fig. 4 A Bedford McMullen self-affine carpet obtained by repeated substitution of the left-hand pattern in itself

The proof depends on careful estimation of the decay of the Fourier transforms of projections of a measure supported by E. The method can be traced back to a study of Bernoulli convolutions by Erdős [12], which Kahane [49] pointed out gave an exceptional set of Hausdorff dimension 0 rather than just Lebesgue measure 0, see [73].

The attractor of an IFS is *self-affine* if (8.1) holds for affine contractions $\{f_1, \ldots, f_k\}$. A plane self-affine set is a *carpet* if the contractions are of the form

$$f_i(x, y) = (a_i x + c_i, b_i y + d_i), \tag{8.4}$$

i.e. affine transformations that leave the horizontal and vertical directions invariant, see Fig. 4. For many self-affine carpets the dimensions of the projections behave well except in directions parallel to the axes.

Theorem 8.5 ([30]) *Let $E \subset \mathbb{R}^2$ be a self-affine carpet in the Bedford-McMullen, Gatzouras-Lalley or Barański class. If the defining IFS is of irrational type, then* $\dim_H \mathrm{proj}_\theta E = \min\{\dim_H E, 1\}$ *for all θ except possibly $\theta = 0$ and $\theta = \frac{1}{2}\pi$.*

For definitions and details of these different classes of carpets, see [30]. The IFS is of *irrational type* if, roughly speaking, $\log a_i / \log b_i$ is irrational for at least one of the f_i in (8.4).

Along similar lines, for an integer $n \geq 2$, let $T_n : [0, 1] \to [0, 1]$ (where 0 and 1 are identified) be given by $T_n(x) = nx \,(\mathrm{mod}\, 1)$. In the 1960s Furstenberg conjectured that if E and F are closed sets invariant under T_2 and T_3 respectively, then $\dim_H \mathrm{proj}_\theta(E \times F)$ should equal $\min\{\dim_H(E \times F), 1\}$ for all θ except possibly $\theta = 0$ and $\theta = \frac{1}{2}\pi$. This was proved by Hochman and Shmerkin [35] along with more general results such as the following.

Theorem 8.6 ([35]) *Let E and F be closed subsets of* $[0, 1]$ *that are invariant under T_m and T_n respectively, where m, n are not powers of the same integer. Then*

$\dim_H \text{proj}_\theta(E \times F) = \min\{\dim_H(E \times F), 1\}$ *for all* θ *except possibly* $\theta = 0$ *and* $\theta = \frac{1}{2}\pi$.

Projection properties of self-affine measures underpin this work and there are measure analogues of these theorems, see [29, 30, 35].

9 Projections of Random Sets

Fractal percolation provides a natural method of generating statistically self-similar fractals, with the same random process determining the form of the fractals at both small and large scales.

Best known is Mandelbrot percolation, based on repeated decomposition of squares into smaller subsquares from which a subset is selected at random. Let D denote the unit square in $\mathbb{R}^2$. Fix an integer $M \geq 2$ and a probability $0 < p < 1$. We divide D into M^2 closed subsquares of side $1/M$ in the obvious way, and retain each subsquare independently with probability p to get a set D_1 formed as a union of the retained subsquares. We repeat this process with the squares of D_1, dividing each into M^2 subsquares of side $1/M^2$ and choosing each with probability p to get a set D_2, and so on. This leads to the random *percolation set* $E = \cap_{k=0}^{\infty} D_k$. If $p > M^{-2}$ then there is a positive probability of non-extinction, i.e. that $E \neq \emptyset$, conditional on which $\dim_H E = 2 + \log p / \log M$ almost surely.

The topological properties of Mandelbrot percolation have been studied extensively, see [11, 17, 78] for surveys. In particular there is a critical probability p_c with $1/M < p_c < 1$ such that if $p > p_c$ then, conditional on non-extinction, E contains many connected components, so projections onto all lines automatically have positive Lebesgue measure. If $p \leq p_c$ the percolation set E is totally disconnected, and Marstrand's theorems provide information on projections of E in almost all directions. However, Rams and Simon [76–78] recently showed using a careful geometrical analysis that, conditional on $E \neq \emptyset$, almost surely the conclusions of Theorem 1.3 hold for all projections.

Theorem 9.1 ([77]) *Let E be the random set obtained by the Mandelbrot percolation process in the plane based on subdivision of squares into M^2 subsquares, each square being retained with probability $p > 1/M^2$. Then, with positive probability $E \neq \emptyset$, conditional on which:*

(i) $\dim_H \text{proj}_\theta E = \min\{\dim_H E, 1\}$ *for all* $\theta \in [0, \pi)$,
(ii) *If* $p > 1/M$ *then for all* $\theta \in [0, \pi)$, $\text{proj}_\theta E$ *contains an interval and in particular* $\mathcal{L}(\text{proj}_\theta E) > 0$.

The natural higher dimensional analogues of this theorem for projections onto all $V \in G(n, m)$ are also valid, see [83]. There are also versions of this result when the squares are selected using certain other probability distributions.

Statistically self-similar subsets of any self-similar set may be constructed using a similar percolation process. Let $\{f_1, \ldots, f_m\}$ be an IFS on $\mathbb{R}^n$ given by (8.2) and let

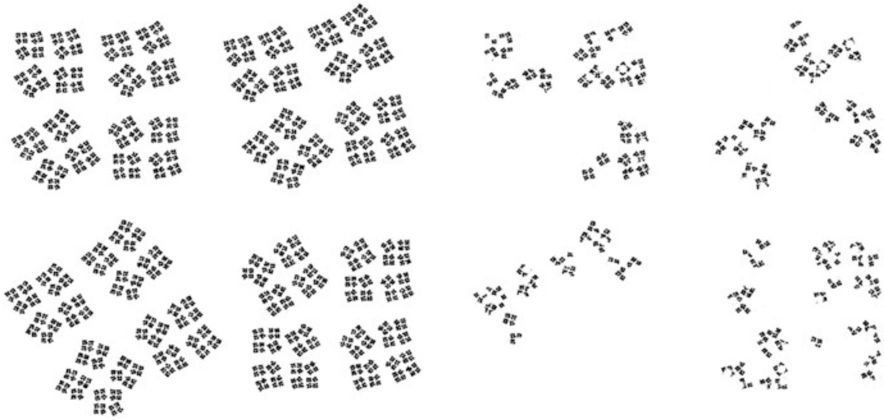


Fig. 5 A self-similar set with dense rotations and a subset obtained by the percolation process

E_0 be its attractor. Percolation on E_0 may be performed by retaining or deleting components of the natural hierarchical construction of E in a random but self-similar manner. Let $0 < p < 1$ and let $D \subset \mathbb{R}^n$ be a non-empty compact set such that $f_i(D) \subset D$ for all i. We select a subfamily of the sets $\{f_1(D), \ldots, f_m(D)\}$ where each $f_i(D)$ is selected independently with probability p and write D_1 for the union of the selected sets. Then, for each selected $f_i(D)$, we choose sets from $\{f_i f_1(D), \ldots, f_i f_m(D)\}$ independently with probability p independently for each i, with the union of these sets comprising D_2. Continuing in this way, we get a nested hierarchy $D \supset D_1 \supset D_2 \supset \cdots$ of random compact sets, where D_k is the union of the components remaining at the kth stage. The random percolation set is $E = \cap_{k=1}^{\infty} D_k$, see Fig. 5. When the underlying IFS has dense rotations, Falconer and Jin [21] extended the ergodic theoretic methods of [35] to random cascade measures to obtain a random analogue of Theorem 8.2.

Theorem 9.2 ([21]) *Let $E_0 \subset \mathbb{R}^n$ be a self-similar set with dense rotation group and let $E \subset E_0$ be the percolation set described above. If $p > 1/m$ there is positive probability that $E \neq \emptyset$, conditional on which,*

$$\dim_{\mathrm{H}} \mathrm{proj}_V E = \min\{\dim_{\mathrm{H}} E, m\} \textit{ for all } V \in G(n, m).$$

More generally, conditional on $E \neq \emptyset$, $\dim_{\mathrm{H}} g(E) = \min\{\dim_{\mathrm{H}} E, m\}$ for all C^1 mappings $g : E \to \mathbb{R}^m$ without singular points.

Recently, Shmerkin and Suomala [82] have introduced a very general theory showing that for a class of random measures, termed spatially independent martingales, very strong results hold for dimensions of projections and sections of the measures, and thus of underlying sets, with the conclusions holding almost surely for projections in all directions or onto all subspaces. Such conclusions are obtained by showing that almost surely the total measures of intersections of the random measures with parameterized deterministic families of measures are absolutely continuous with respect to the parameter. Spatially independent measures include measures based on fractal percolation, random cascades and random cutout models.

10 Further Variations and Applications of Projections

This discussion has covered just a few of the numerous results which may be traced back to Marstrand's pioneering work. We end with an even briefer mention of some further applications, with one or two references indicating where further information may be found.

Visible parts of sets The *visible part* $\mathrm{Vis}_\theta E$ of a compact set $E \subset \mathbb{R}^2$ from direction θ is the set of $x \in E$ such that the half-line from x in direction θ intersects E in the single point x; thus $\mathrm{Vis}_\theta E$ may be thought of as the part of E that can be 'seen from infinity' in direction θ. It is immediate from Marstrand's Theorem 1.3 that, for almost all θ,

$$\dim_{\mathrm{H}} \mathrm{Vis}_\theta E = \dim_{\mathrm{H}} E \text{ if } \dim_{\mathrm{H}} E \le 1 \quad \text{and} \quad \dim_{\mathrm{H}} \mathrm{Vis}_\theta E \ge 1 \text{ if } \dim_{\mathrm{H}} E \ge 1.$$

It has been conjectured that if $\dim_{\mathrm{H}} E \ge 1$ then $\dim_{\mathrm{H}} \mathrm{Vis}_\theta E = 1$ for almost all θ, but so far this has only been established for rather specific classes of E. The conjecture is easily verified if E is the graph of a function (the only exceptional direction being perpendicular to the x-axis), see [47]. It is also known for quasi-circles [47] and for Mandelbrot percolation sets [1]. For self-similar sets, it holds if E is connected and the rotation group is finite [1], and also if E satisfies the open set condition for a convex open set such that $\mathrm{proj}_\theta E$ is an interval for all θ [18] (in this case E need not be connected). The analogous conjecture in higher dimensions, that the dimension of the visible part of a set $E \subset \mathbb{R}^n$ equals $\min\{\dim_{\mathrm{H}} E, n-1\}$, is also unresolved if $\dim_{\mathrm{H}} E > n-1$.

Projections in infinite dimensional spaces Infinite-dimensional dynamical systems may have finite dimensional attractors. When they are studied experimentally what is observed is essentially a projection or 'embedding' of the attractor into Euclidean space and infinite-dimensional versions of the projection theorems can relate these projections to the original attractor. Let E be a compact subset of a Banach space X with box-counting (or Minkowski) dimension d. Hunt and Kaloshin [41] show that for almost every projection or bounded linear function $\pi : X \to \mathbb{R}^m$ such that $m > 2d$,

$$\frac{m-2d}{m(1+d)} \dim_{\mathrm{H}} E \le \dim_{\mathrm{H}} \pi(E) \le \dim_{\mathrm{H}} E.$$

Here 'almost every' is interpreted in the sense of *prevalence*, which is a measure-theoretic way of defining sparse and full sets for infinite-dimensional spaces. The book by Robinson [79] provides a recent treatment of this important area.

Projections in Heisenberg groups The Heisenberg group $\mathbb{H}^n$ is the connected and simply connected nilpotent Lie group of step 2 and dimension $2n + 1$ with 1-dimensional center, which may be identified topologically with $\mathbb{R}^{2n+1}$. However, the Heisenberg metric d_H, which is invariant under the group action, is very different from the Euclidean metric, and in particular the Hausdorff dimension of subsets of $\mathbb{H}^n$ depends on which metric is used in the definition. Despite the lack of isotropy, there is enough geometric structure to enable families of projections to be defined, and it is possible to get bounds for the dimensions of certain projections of a Borel set E in terms of the dimension of E, where the dimensions are defined with respect to d_H, see [2, 3, 61].

Sections of sets Dimensions of sections or slices of sets, which go hand in hand with dimensions of projections, also featured in Marstrand's 1954 paper [56]. He showed essentially that, if $E \subset \mathbb{R}^2$ is a Borel or analytic set of Hausdorff dimension $s > 1$, then for almost all directions θ, $\dim_{\mathrm{H}} \mathrm{proj}_\theta^{-1} x \leq s - 1$ for almost all $x \in V_\theta$, with equality for a set of $x \in V_\theta$ of positive Lebesgue measure. Here $\mathrm{proj}_\theta : \mathbb{R}^2 \to V_\theta$ is orthogonal projection onto V_θ, the line in direction θ. The natural higher dimensional analogues were obtained by Mattila [57, 58, 60] using potential theoretic arguments. Most of the aspects discussed above for projections have been investigated for sections, including packing dimensions [23, 48], exceptional directions [70], self-similar sets [22, 31] and fractal percolation sets [22, 82].

Projections of measures For μ a Borel measure on $\mathbb{R}^n$ with compact support such that $0 < \mu(\mathbb{R}^n) < \infty$, the projection $\mathrm{proj}_V\mu$ of μ onto a subspace $V \in G(n, m)$ is defined in the natural way, that is by $(\mathrm{proj}_V\mu)(A) = \mu\{x \in \mathbb{R}^n : \mathrm{proj}_V(x) \in A\}$ for Borel sets A or equivalently by $\int f(t)d(\mathrm{proj}_V\mu)(t) = \int f(\mathrm{proj}_V(x))d\mu(x)$ for continuous f. The support of $\mathrm{proj}_V\mu$ is the projection onto V of the support of μ, so it is not surprising that many of the results for projection of sets have analogues for projection of measures. Indeed many projection results for sets are obtained by putting a suitable measure on the set and examining projections of the measure, as in Kaufman's proof in Sect. 2. There are many ways of quantifying the fine structure of measures, and the way these behave under projections have been investigated in many cases.

For example, the *lower pointwise* or *local dimension* of a Borel probability measure μ on $\mathbb{R}^n$ at $x \in \mathbb{R}^n$ is given by $\underline{\dim}_\mu(x) = \underline{\lim}_{r\to 0} \log \mu(B(x, r))/ \log r$, with a corresponding definition taking the upper limit for the *upper pointwise dimension*. Then, for almost all every subspace $V \in G(n, m)$ and μ-almost all $x \in \mathbb{R}^n$,

$$\underline{\dim}_\mu(\mathrm{proj}_V x) = \min\{\underline{\dim}_\mu(x), m\} \quad \text{and} \quad \overline{\dim}_\mu(\mathrm{proj}_V x) = \min\{\overline{\dim}_\mu(x), m\},$$

see [24, 35, 39, 40, 86]. The (*lower*) *Hausdorff dimension of a measure* μ is defined as $\dim_{\mathrm{H}} \mu = \inf\{\dim_{\mathrm{H}} A : \mu(A) > 0\}$. It follows easily from the projection properties of pointwise dimension that, for almost every V,

$$\dim_{\mathrm{H}}(\mathrm{proj}_V \mu) = \min\{\dim_{\mathrm{H}} \mu(x), m\}.$$

The L_q-dimensions of projections are examined in [40], for the multifractal spectrum see [4, 66, 67], and for packing dimension aspects see [23].

For a special case of projection of measures, let M be a compact Riemann surface and $\mathrm{proj} : T^1 M \to M$ be the natural projection from the unit tangent bundle $T^1 M$ to M. Let μ be a probability measure on $T^1 M$ that is invariant under the geodesic flow on $T^1 M$. Ledrappier and Lindenstrauss [54] showed that if $\dim_{\mathrm{H}} \mu \le 2$ then $\dim_{\mathrm{H}} \mathrm{proj}\mu = \dim_{\mathrm{H}} \mu$, and if $\dim_{\mathrm{H}} \mu > 2$ then $\mathrm{proj}\mu$ is absolutely continuous see also [36, 37]. However, the analogous conclusion fails if the base manifold has dimension 3 or more, see [46].

Other results, such as Theorems 8.2 & 9.2 have natural measure analogues.

11 Conclusion

If this article does nothing else, it should demonstrate just how much of fractal geometry has its roots in Marstrand's 1954 paper. If further evidence is needed, there are hundreds of citations of the paper in MathSciNet and Google Scholar, despite these indexes only including relatively recent references.

This survey of projection results has been brief and far from exhaustive and there are many more related papers. For a both broader and more detailed coverage of various aspects of projections, the books by Falconer [16, 17] and Mattila [58, 62] and the survey articles by Mattila [59–61] may be helpful.

References

1. I. Arhosalo, E. Järvenpää, M. Järvenpää, M. Rams, P. Shmerkin, Visible parts of fractal percolation. Proc. Edin. Math. Soc. (2), **55**, 311–331 (2012)
2. Z.M. Balogh, D.E. Cartagena, K. Fässler, P. Mattila, J.T. Tyson, The effect of projections on dimension in the Heisenberg group. Rev. Mat. Iberoam. **29**, 381–432 (2013)
3. Z.M. Balogh, K. Fässler, P. Mattila, J.T. Tyson, Projection and slicing theorems in Heisenberg groups. Adv. Math. **231**, 569–604 (2012)
4. J. Barral, I. Bhouri, Multifractal analysis for projections of Gibbs and related measures. Ergod. Theory Dyn. Syst. **31**, 673–701 (2011)
5. A.S. Besicovitch, On the fundamental geometrical properties of linearly measurable plane sets of points, I. Math. Ann. **98**, 422–464 (1927)
6. A.S. Besicovitch, On the fundamental geometrical properties of linearly measurable plane sets of points, II. Math. Ann. **115**, 296–329 (1938)
7. A.S. Besicovitch, On the fundamental geometrical properties of linearly measurable plane sets of points, III. Math. Ann. **116**, 349–357 (1939)

8. J. Bougain, On the Erdős-Volkmann and Katz-Tao ring conjectures. Geom. Funct. Anal. **13**, 334–365 (2003)
9. R.O. Davies, Subsets of finite measure in analytic sets. Indag. Math. **14**, 488–489 (1952)
10. R.O. Davies, On accessibility of plane sets and differentiation of functions of two real variables. Proc. Camb. Philos. Soc. **48**, 215–232 (1952)
11. M. Dekking, Random Cantor sets and their projections, in *Fractal Geometry and Stochastics IV*. Progress in Probability, vol. 61 (Birkhäuser Verlag, Basel, 2009), pp. 269–284
12. P. Erdős, On a family of symmetric Bernoulli convolutions. Am. J. Math. **61**, 974–976 (1935)
13. K.I. Eroğlu, On planar self-similar sets with a dense set of rotations. Ann. Acad. Sci. Fenn. A Math. **32**, 409–424 (2007)
14. K.J. Falconer, Hausdorff dimension and the exceptional set of projections. Mathematika **29**, 109–115 (1982)
15. K.J. Falconer, Sets with prescribed projections and Nikodym sets. Proc. Lond. Math. Soc. (3) **53**, 48–64 (1982)
16. K.J. Falconer, *The Geometry of Fractal Sets* (Cambridge University Press, Cambridge, 1985)
17. K.J. Falconer, *Fractal Geometry: Mathematical Foundations and Applications*, 3rd edn. (Wiley, Hoboken, 2014)
18. K.J. Falconer, J.M. Fraser, The visible part of plane self-similar sets. Proc. Am. Math. Soc. **141**, 269–278 (2013)
19. K.J. Falconer, J.D. Howroyd, Projection theorems for box and packing dimensions. Math. Proc. Camb. Philos. Soc. **119**, 287–295 (1996)
20. K.J. Falconer, J.D. Howroyd, Packing dimensions of projections and dimension profiles. Math. Proc. Camb. Philos. Soc. **121**, 269–286 (1997)
21. K.J. Falconer, X. Jin, Exact dimensionality and projections of random self-similar measures and sets. J. Lond. Math. Soc. (2) **90**, 388–412 (2014)
22. K.J. Falconer, X. Jin, Dimension conservation for self-similar sets and fractal percolation IMRN, Advanced Access, doi:10.1093/imrn/rnv103 (2015). arXiv:1409.1882
23. K.J. Falconer, P. Mattila, The packing dimension of projections and sections of measures. Math. Proc. Camb. Philos. Soc. **119**, 695–713 (1996)
24. K.J. Falconer, T.C. O'Neil, Convolutions and the geometry of multifractal measures. Math. Nachr. **204**, 61–82 (1999)
25. A. Farkas, Projections of self-similar sets with no separation condition to appear, Isr. J. Math. (2015). arXiv:1307.2841
26. K. Fässler, T. Orponen, On restricted families of projections in $\mathbb{R}^3$. Proc. Lond. Math. Soc. **109**, 353–381 (2014)
27. H. Federer, The (ϕ, k) rectifiable subsets of n space. Trans. Am. Math. Soc. **62**, 114–192 (1947)
28. H. Federer, *Geometric Measure Theory* (Springer, Berlin/Heidelberg/New York, 1969). pbk. reprint 1996
29. A. Ferguson, J. Fraser, T. Sahlsten, Scaling scenery of $(\times m, \times n)$ invariant measures. Adv. Math. **268**, 564–602 (2015)
30. A. Ferguson, T. Jordan, P. Shmerkin, The Hausdorff dimension of the projections of self-affine carpets. Fund. Math. **209**, 193–213 (2010)
31. H. Furstenberg, Ergodic fractal measures and dimension conservation. Ergod. Theory Dyn. Syst. **28**, 405–422 (2008)
32. H. Furstenberg, *Ergodic Theory and Fractal Geometry* (American Mathematical Society & Conference Board of Mathematical Sciences, Providence, 2014)
33. M. Hochman, Dynamics on fractals and fractal distributions (2013). arXiv:1008.3731v2
34. M. Hochman, On self-similar sets with overlaps and inverse theorems for entropy. Ann. Math.(2) **180**, 773–822 (2014)
35. M. Hochman, P. Shmerkin, Local entropy averages and projections of fractal measures. Ann. Math.(2) **175**, 1001–1059 (2012)
36. R. Hovila, E. Järvanpää, M. Järvanpää, F. Ledrappier, Singularity of projections of 2-dimensional measures invariant under the geodesic flow. Commun. Math. Phys. **312**, 127–136 (2012)

37. R. Hovila, E. Järvenpää, M. Järvenpää, F. Ledrappier, Besicovitch-Federer projection theorem and geodesic flows on Riemann surfaces. Geom. Dedicata **161**, 51–61 (2012)
38. J.D. Howroyd, Box and packing dimensions of projections and dimension profiles. Math. Proc. Camb. Philos. Soc. **130**, 135–160 (2001)
39. X. Hu, S.J. Taylor, Fractal properties of products and projections of measures in $\mathbb{R}^d$. Math. Proc. Camb. Philos. Soc. **115**, 527–544 (1994)
40. B.R. Hunt, Y. Kaloshin, How projections affect the dimension spectrum of fractal measures. Nonlinearity **10**, 1031–1046 (1997)
41. B.R. Hunt, Y. Kaloshin, Regularity of embeddings of infinite-dimensional fractal sets into finite-dimensional spaces. Nonlinearity **12**, 1263–1275 (2008)
42. J.E. Hutchinson, Fractals and self-similarity. Indiana Univ. Math. J. **30**, 713–747 (1981)
43. M. Järvenpää, On the upper Minkowski dimension, the packing dimension, and othogonal projections. Ann. Acad. Sci. Fenn. A Diss. **99**, 34 pp (1994)
44. E. Järvenpää, M. Järvenpää, T. Keleti, Hausdorff dimension and non-degenerate families of projections. J. Geom. Anal. **24**, 2020–2034 (2014)
45. E. Järvenpää, M. Järvenpää, F. Ledrappier, M. Leikas, One-dimensional families of projections. Nonlinearity **21**, 453–463 (2008)
46. E. Järvenpää, M. Järvenpää, M. Leikas, (Non)regularity of projections of measures invariant under geodesic flow. Commun. Math. Phys. **254**, 695–717 (2005)
47. E. Järvenpää, M. Järvenpää, P. MacManus, T.C. O'Neil, Visible parts and dimensions. Nonlinearity **16**, 803–818 (2003)
48. M. Järvenpää, P. Mattila, Hausdorff and packing dimensions and sections of measures. Mathematika **45**, 55–77 (1998)
49. J.-P. Kahane, Sur la distribution de certaines séries aléatoires, in *Colloque de Théorie des Nombres, Univ. Bordeaux, 1969*. Bull. Soc. Math. France, Mém., vol. 25 (Soc. Math. France, Paris, 1971), pp. 119–122
50. R. Kaufman, On Hausdorff dimension of projections. Mathematika **15**, 153–155 (1968)
51. R. Kaufman, P. Mattila, Hausdorff dimension and exceptional sets of linear transformations. Ann. Acad. Sci. Fenn. A Math. **1**, 387–392 (1975)
52. R. Kenyon, Projecting the one-dimensional Sierpinski gasket. Isr. J. Math. **97**, 221–238 (1997)
53. D. Khoshnevisan, Y. Xiao, Packing-dimension profiles and fractional Brownian motion. Math. Proc. Camb. Philos. Soc. **145**, 145–213 (2008)
54. F. Ledrappier, E. Lindenstrauss, On the projections of measures invariant under the geodesic flow. IMRN **9**, 511–526 (2003)
55. M. Leikas, Packing dimensions, transversal mappings and geodesic flows. Ann. Acad. Sci. Fenn. A Math. **29**, 489–500 (2004)
56. J.M. Marstrand, Some fundamental geometrical properties of plane sets of fractional dimensions. Proc. Lond. Math. Soc.(3) **4**, 257–302 (1954)
57. P. Mattila, Hausdorff dimension, orthogonal projections and intersections with planes. Ann. Acad. Sci. Fenn. A Math. **1**, 227–244 (1975)
58. P. Mattila, *Geometry of Sets and Measures in Euclidean Spaces* (Cambridge University Press, Cambridge, 1995)
59. P. Mattila, Hausdorff dimension, projections, and the Fourier transform. Publ. Mat. **48**, 3–48 (2004)
60. P. Mattila, Marstrand's theorems, in *All That Math, Portraits of Mathematicians as Young Readers*, ed. by A. Córdoba, J.L. Férnandez, P. Férnandez (Revista Matemática Iberoamericana, Madrid, 2011), pp. 283–301
61. P. Mattila, Recent progress on dimensions of projections, in *Geometry and Analysis of Fractals*, ed. by D.-J. Feng, K.-S. Lau. Springer Proceedings in Mathematics & Statistics, vol. 88 (Springer, Berlin/Heidelberg, 2014), pp. 283–301
62. P. Mattila, *Fourier Analysis and Hausdorff Dimension* (Cambridge University Press, Cambridge, 2015)
63. D.M. Oberlin, Restricted Radon transforms and projections of planar sets. Can. Math. Bull. **55**, 815–820 (2012)

64. D.M. Oberlin, Exceptional sets of projections, unions of *k*-planes and associated transforms. Isr. J. Math. **202**, 331–342 (2014)
65. D.M. Oberlin, R. Oberlin, Application of a Fourier restriction theorem to certain families of projections in $\mathbb{R}^3$. J. Geom. Anal. (to appear) arXiv:1307.5039 (2013)
66. L. Olsen, Multifractal geometry, in *Fractal Geometry and Stochastics, II, Greifswald/Koserow, 1998*. Progress in Probability, vol. 46 (Birkhäuser, Basel, 2000), pp. 3–37
67. T.C. O'Neil, The multifractal spectra of projected measures in Euclidean spaces. Chaos Solitons Fractals **11**, 901–921 (2000)
68. T. Orponen, On the packing dimension and category of exceptional sets of orthogonal projections Ann. Mat. Pura Appl. (4) **194**, 843–880 (2015)
69. T. Orponen, On the distance sets of self-similar sets. Nonlinearity **25**, 1919–1929 (2012)
70. T. Orponen, Slicing sets and measures, and the dimension of exceptional parameters. J. Geom. Anal. **24**, 47–80 (2012)
71. T. Orponen, Hausdorff dimension estimates for restricted families of projections in $\mathbb{R}^3$ Adv. Math. **275**, 147–183 (2015)
72. Y. Peres, B. Schlag, Smoothness of projections, Bernoulli convolutions, and the dimension of exceptions. Duke Math. J. **102**, 193–251 (2000)
73. Y. Peres, W. Schlag, B. Solomyak, Sixty years of Bernoulli convolutions, in *Fractal Geometry and Stochastics, II, Greifswald/Koserow, 1998*. Progress in Probability, vol. 46 (Birkhäuser, Basel, 2000), pp. 39–65
74. Y. Peres, P. Shmerkin, Resonance between Cantor sets. Ergod. Theory Dyn. Syst. **29**, 201–221 (2009)
75. M. Pollicott, K. Simon, The Hausdorff dimension of projections of λ-expansions with deleted digits. Trans. Am. Math. Soc. **347**, 967–983 (1995)
76. M. Rams, K. Simon, Projections of fractal percolations. Ergod. Theory Dyn. Syst. **35**, 530–545 (2015)
77. M. Rams, K. Simon, The dimension of projections of fractal percolations. J. Stat. Phys. **154**, 633–655 (2014)
78. M. Rams, K. Simon, The geometry of fractal percolation, in *Geometry and Analysis of Fractals*, ed. by D.-J. Feng, K.-S. Lau. Springer Proceedings in Mathematics & Statistics, vol. 88 (Springer, Berlin/Heidelberg, 2014), pp 303–324
79. J.C. Robinson, *Dimensions, Embeddings, and Attractors* (Cambridge University Press, Cambridge, 2010)
80. P. Shmerkin, Projections of self-similar and related fractals: a survey of recent developments, in *Fractal Geometry and Stochastics 5 (Tabarz 2014)* (Birkhäuser, Basel, 2015), pp 53–74
81. P. Shmerkin, B. Solomyak, Absolute continuity of self-similar measures, their projections and convolutions (2014). arXiv:1406.0204
82. P. Shmerkin, V. Suomala, Spatially independent martingales, intersections, and applications (2015). arXiv:1409.6707v4
83. K. Simon, L. Vágó, Projections of Mandelbrot percolation in higher dimensions. Ergod. Theory Dyn. Syst. (2014, to appear). arXiv:1407.2225
84. C. Tricot, Two definitions of fractional dimension. Math. Proc. Camb. Philos. Soc. **91**, 57–74 (1982)
85. Y. Xiao, Packing dimension of the image of fractional Brownian motion. Stat. Probab. Lett. **333**, 379–387 (1997)
86. M. Zähle, The average fractal dimension and projections of measures and sets in $\mathbb{R}^n$. Fractals **3**, 747 (1995)

Scenery Flow, Conical Densities, and Rectifiability

Antti Käenmäki

Abstract We present an application of the recently developed ergodic theoretic machinery on scenery flows to a classical geometric measure theoretic problem in Euclidean spaces. We also review the enhancements to the theory required in our work. Our main result is a sharp version of the conical density theorem, which we reduce to a question on rectifiability.

Keywords Scenery flow • Fractal distributions • Conical densities • Rectifiability

Mathematics Subject Classification (2000). Primary 28A80; Secondary 37A10, 28A75, 28A33

1 Introduction

We survey a recent advance in the study of scenery flows and show how it can be applied in a classical question in geometric measure theory which a priori does not involve any dynamics. The reader is prompted to recall the expository article of Fisher [8] where it was discussed how the scenery flow is linked to rescaling on several well-studied structures, such as geodesic flows, Brownian motion, and Julia sets. The purpose of this note is to continue that line of introduction.

The idea behind the scenery flow has been examined in many occasions. Authors have considered the scenery flow for specific sets and measures arising from dynamics; see e.g. [1–3, 7, 21]. Abstract scenery flows have also been studied with a view on applications to special sets and measures, again arising from dynamics or arithmetic; see e.g. [11–13]. The main innovation of the recent article by Käenmäki, Sahlsten, and Shmerkin [15] is to employ the general theory initiated by

A. Käenmäki (✉)
Department of Mathematics and Statistics, P.O. Box 35 (MaD), FI-40014 University of Jyväskylä, Finland
e-mail: antti.kaenmaki@jyu.fi

C. Bandt et al. (eds.), *Fractal Geometry and Stochastics V*, Progress in Probability 70, DOI 10.1007/978-3-319-18660-3_2

Furstenberg [10], greatly developed by Hochman [12] and extended by Käenmäki, Sahlsten, and Shmerkin [16], to classical problems in geometric measure theory.

One of the most fundamental concepts of geometric measure theory is that of rectifiability. It is a measure-theoretical notion for smoothness and to a great extend, geometric measure theory is about studying rectifiable sets. The foundations of geometric measure theory were laid by Besicovitch [4, 5]. For various characterizations and properties of rectifiability the reader is referred to the book of Mattila [18]. In conical density results, the idea is to examine how a measure is distributed in small balls. Finding conditions that guarantee the measure to be effectively spread out in different directions is a classical question going back to Besicovitch [6] and Marstrand [17]. For an account of the development on conical density results the reader is referred to the survey of Käenmäki [14].

The scenery flow is a well-suited tool to address problems concerning conical densities. The cones in question do not change under magnification and this allows to pass information between the original measure and its tangential structure. In fact, we will see that there is an intimate connection between rectifiability and conical densities.

This exposition comes in two parts. In the first part, we review dynamical aspects of the scenery flow and in the second part, we focus on geometric measure theory.

2 Dynamics of the Scenery Flow

Let $(X, \mathcal{B}, P)$ be a probability space. We shall assume that X is a metric space and $\mathcal{B}$ is the Borel σ-algebra on X. Write $\mathbb{R}_+ = [0, \infty)$. A (one-sided) *flow* is a family $(F_t)_{t\in\mathbb{R}_+}$ of measurable maps $F_t\colon X \to X$ for which

$$F_{t+t'} = F_t \circ F_{t'}, \quad t, t' \in \mathbb{R}_+.$$

In other words, $(F_t)_{t\in\mathbb{R}_+}$ is an additive $\mathbb{R}_+$ action on X. We also assume that $(x, t) \mapsto F_t(x)$ is measurable.

We say that a set $A \in \mathcal{B}$ is F_t *invariant* if $P(F_t^{-1}A \triangle A) = 0$ for all $t \geq 0$. If $F_t P = P$ for all $t \geq 0$, then we say that P is F_t *invariant*. In this case, we call $(X, \mathcal{B}, P, (F_t)_{t\in\mathbb{R}_+})$ a *measure preserving flow*. Furthermore, a measure preserving flow is *ergodic*, if for all $t \geq 0$ the measure P is ergodic with respect to the transformation $F_t\colon X \to X$, that is, for all F_t invariant sets $A \in \mathcal{B}$ we have $P(A) \in \{0, 1\}$.

Theorem 2.1 (Birkhoff ergodic theorem) *If $(X, \mathcal{B}, P, (F_t)_{t\in\mathbb{R}_+})$ is an ergodic measure preserving flow, then for a P integrable function $f\colon X \to \mathbb{R}$ we have*

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T f(F_t x)\, \mathrm{d}t = \int f \,\mathrm{d}P$$

for P-almost all $x \in X$.

We write $\omega \sim P$ to indicate that ω is chosen randomly according to the measure P.

Theorem 2.2 (Ergodic decomposition) *Any F_t invariant measure P can be decomposed into ergodic components P_ω, $\omega \sim P$, such that*

$$P = \int P_\omega \, \mathrm{d}P(\omega).$$

This decomposition is unique up to P measure zero sets.

Let us next define the scenery flow. We equip $\mathbb{R}^d$ with the usual Euclidean norm and the induced metric. Denote the closed unit ball by B_1. Let $\mathcal{M}_1 := \mathcal{P}(B_1)$ be the collection of all Borel probability measures on B_1 and $\mathcal{M}_1^* := \{\mu \in \mathcal{M}_1 : 0 \in \mathrm{spt}(\mu)\}$. Here $\mathrm{spt}(\mu)$ is the support of μ. To avoid any confusion, measures on measures will be called *distributions*. We define the *magnification* $S_t\mu$ of $\mu \in \mathcal{M}_1^*$ at 0 by setting

$$S_t\mu(A) := \frac{\mu(e^{-t}A)}{\mu(B(0, e^{-t}))}, \quad A \subset B_1.$$

In other words, the measure $S_t\mu$ is obtained by scaling $\mu|_{B(0,e^{-t})}$ into the unit ball and normalizing. Due to the exponential scaling, $(S_t)_{t \in \mathbb{R}_+}$ is a flow in the space $\mathcal{M}_1^*$ and we call it the *scenery flow* at 0. An S_t invariant distribution P on $\mathcal{M}_1^*$ is called *scale invariant*. Although the action S_t is discontinuous (at measures μ with $\mu(\partial B(0,r)) > 0$ for some $0 < r < 1$) and the set $\mathcal{M}_1^* \subset \mathcal{M}_1$ is not closed, we shall witness that the scenery flow behaves in a very similar way to a continuous flow on a compact metric space.

With the scenery flow we are now able to define tangent measures and distributions. Let μ be a Radon measure and $x \in \mathrm{spt}(\mu)$. We want to consider the scaling dynamics when magnifying around x. Let $T_x\mu(A) := \mu(A + x)$ and define $\mu_{x,t} := S_t(T_x\mu)$. Then the one-parameter family $(\mu_{x,t})_{t \in \mathbb{R}_+}$ is called the *scenery flow* at x. Accumulation points of this scenery in $\mathcal{M}_1$ will be called *tangent measures* of μ at x and the family of tangent measures of μ at x is denoted by $\mathrm{Tan}(\mu, x) \subset \mathcal{M}_1$. However, we are not interested in a single tangent measure, but the whole statistics of the scenery $\mu_{x,t}$ as $t \to \infty$. We remark that we have slightly deviated from Preiss' original definition of tangent measures, which corresponds to taking weak limits of unrestricted blow-ups; see [20].

Definition 2.3 (Tangent distributions) A *tangent distribution* of μ at $x \in \mathrm{spt}(\mu)$ is any weak limit of

$$\langle \mu \rangle_{x,T} := \frac{1}{T} \int_0^T \delta_{\mu_{x,t}} \, \mathrm{d}t$$

as $T \to \infty$. The family of tangent distributions of μ at x is denoted by $\mathcal{TD}(\mu, x) \subset \mathcal{P}(\mathcal{M}_1^*)$.

If the limit above is unique, then, intuitively, it means that the collection of views $\mu_{x,t}$ will have well-defined statistics when zooming into smaller and smaller neighbourhoods of x. The integration above makes sense since we are on a convex subset of a topological linear space. We emphasize that tangent distributions are measures on measures. Notice that the set $\mathcal{TD}(\mu, x)$ is non-empty and compact at $x \in \operatorname{spt}(\mu)$. Moreover, the support of each $P \in \mathcal{TD}(\mu, x)$ is contained in $\operatorname{Tan}(\mu, x)$.

According to Preiss' well-known principle, tangent measures to tangent measures are tangent measures; see [20, Theorem 2.12]. We shall define an analogous condition for distributions. We say that a distribution P on $\mathcal{M}_1$ is *quasi-Palm* if for any Borel set $\mathcal{A} \subset \mathcal{M}_1$ with $P(\mathcal{A}) = 1$ it holds that for P-almost every $\nu \in \mathcal{A}$ and for ν-almost every $z \in \mathbb{R}^d$ there exists $t_z > 0$ such that for $t \geq t_z$ we have $B(z, e^{-t}) \subset B_1$ and

$$\nu_{z,t} \in \mathcal{A}.$$

This version of the quasi-Palm property actually requires that the unit sphere of the norm is a C^1 manifold and does not contain line segments; see [15, Lemma 3.23]. The Euclidean norm we use of course satisfies this requirement. If we were considering unrestricted blow-ups, then the requirement for $B(z, e^{-t})$ to be contained in B_1 could be dropped. Roughly speaking, the quasi-Palm property guarantees that the null sets of the distributions are invariant under translations to a typical point of the measure.

Definition 2.4 (Fractal distributions) A distribution P on $\mathcal{M}_1$ is a *fractal distribution* if it is scale invariant and quasi-Palm. A fractal distribution is an *ergodic fractal distribution* if it is ergodic with respect to S_t.

It follows from the Besicovitch density point theorem that ergodic components of a fractal distribution are ergodic fractal distributions; see [12, Theorem 1.3].

A general principle is that tangent objects enjoy some kind of spatial invariance. For tangent distributions, a very powerful formulation of this principle is the following theorem of Hochman [12, Theorem 1.7]. The result is analogous to a similar phenomenon discovered by Mörters and Preiss [19, Theorem 1].

Theorem 2.5 *For any Radon measure μ and μ-almost every x, all tangent distributions of μ at x are fractal distributions.*

Notice that as the action S_t is discontinuous, even the scale invariance of tangent distributions or the fact that they are supported on $\mathcal{M}_1^*$ are not immediate, though they are perhaps expected. The most interesting part in the above theorem is that a typical tangent distribution satisfies the quasi-Palm property.

Hochman's result is proved by using CP processes which are Markov processes on the dyadic scaling sceneries of a measure introduced by Furstenberg [9, 10]. Let $\mathcal{D}$ be a partition of $[-1, 1]^d$ into 2^d cubes of side length 1. Given $x \in [-1, 1]^d$, let $D(x)$ be the only element of $\mathcal{D}$ containing it. If $D \in \mathcal{D}$, then we write T_D for the orientation preserving homothety mapping from $\overline{D}$ onto $[-1, 1]^d$. Define the *CP magnification* M on $\Omega := \mathcal{P}([-1, 1]^d) \times [-1, 1]^d$ by setting

$$M(\mu, x) := \big(T_{D(x)}\mu/\mu(D(x)), T_{D(x)}(x)\big).$$

This is well-defined whenever $\mu(D(x)) > 0$. Note that, since zooming in is done dyadically, it is important to keep track of the orbit of the point that is being zoomed upon. A distribution Q on Ω is *adapted* if there is a disintegration

$$\int f(\nu, x)\, \mathrm{d}Q(\nu, x) = \iint f(\nu, x)\, \mathrm{d}\nu(x)\, \mathrm{d}\overline{Q}(\nu)$$

for all $f \in C(\Omega)$. Here $\overline{Q}$ is the projection of Q onto the measure component. In other words, Q is adapted if choosing a pair (μ, x) according to Q can be done in a two-step process, by first choosing μ according to $\overline{Q}$ and then choosing x according to μ. A distribution on Ω is a *CP distribution* if it is M invariant and adapted.

The *micromeasure distribution* of μ at $x \in \operatorname{spt}(\mu)$ is any weak limit of

$$\langle \mu, x\rangle_N := \frac{1}{N}\sum_{k=0}^{N-1} \delta_{M^k(\mu,x)}.$$

By compactness of $\mathcal{P}(\Omega)$, the family of micromeasure distributions is non-empty and compact, and by [12, Proposition 5.4], each micromeasure distribution is adapted. Furthermore, if the *intensity measure* of a micromeasure distribution Q defined by

$$[Q](A) := \int \mu(A)\, \mathrm{d}\overline{Q}(\mu), \quad A \subset [-1, 1]^d,$$

is the normalized Lebesgue measure, then Q is M invariant. By adaptedness, this is the case for any weak limit of $\langle \mu + z, x + z\rangle_N$ for Lebesgue almost all $z \in [-1/2, 1/2]^d$; see [12, Proposition 5.5(2)]. In other words, by slightly adjusting the dyadic grid, a micromeasure distribution can be seen to be a CP distribution. The family of CP distributions having Lebesgue intensity is compact; see [16, Lemma 3.4].

If Q is a CP distribution, then the system (Ω, M, Q) is a stationary one-sided process $(\xi_n)_{n\in\mathbb{N}}$ with $\xi_1 \sim Q$ and $M\xi_n = \xi_{n+1}$. Considering its two-sided extension, we see that there exists a natural extension $\hat{Q}$ supported on the Cartesian product of all Radon measures and $[-1, 1]^d$. A *centering* of $\hat{Q}$ is a push-down of the suspension flow of $\hat{Q}$ under the unrestricted magnification of μ at x. For a precise definition, see [12, Definition 1.13]. By [12, Theorem 1.14], a centering of $\hat{Q}$ is an unrestricted fractal distribution. We remark that [12] and [16] use L^∞ norm to allow an easier link between CP processes and fractal distributions. By [16, Appendix A], the results are independent of the choice of the norm and hence, our use of the Euclidean norm is justified.

Relying on the above, we are now able to give an outline for the proof of Theorem 2.5. If $P = \lim_{k\to\infty}\langle\mu\rangle_{x,N_k}$ is a tangent distribution, then, passing to a subsequence, define a micromeasure distribution $Q = \lim_{i\to\infty}\langle\mu, x\rangle_{N_{k(i)}}$. Slightly adjusting the dyadic grid, we see that Q is a CP distribution with Lebesgue intensity. Thus, by [12, Proposition 5.5(3)], P is the restriction of the centering of $\hat{Q}$ and hence, P is a fractal distribution.

Although fractal distributions are defined in terms of seemingly strong geometric properties, the family of fractal distributions is in fact very robust. The following theorem is due to Käenmäki, Sahlsten, and Shmerkin [16, Theorem A].

Theorem 2.6 *The family of fractal distributions is compact.*

The result may appear rather surprising since the scenery flow is not continuous, its support is not closed, and, more significantly, the quasi-Palm property is not a closed property. The proof of this result is also based on the interplay between fractal distributions and CP processes. We have already seen that each CP distribution defines a fractal distribution. The converse is also true. Let us first assume that P is an ergodic fractal distribution. If f is a continuous function defined on $\mathcal{P}(\mathcal{M}_1)$, then, by the Birkhoff ergodic theorem, we have

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T f(S_t\mu)\,\mathrm{d}t = \int f\,\mathrm{d}P$$

for P-almost all μ. Considering a countable dense set of continuous functions f and applying the quasi-Palm property, it follows that

$$\lim_{T\to\infty}\langle\mu\rangle_{x,T} = P \tag{2.1}$$

for P-almost all μ and for μ-almost all x; see [12, Theorem 3.9]. As we already have seen, any tangent distribution can be expressed as the restriction of the centering of an extended CP distribution having Lebesgue intensity. Thus, by (2.1), the same holds for ergodic fractal distributions. Relying on the ergodic decomposition, this observation can be extended to non-ergodic fractal distributions; see [12, Theorem 1.15]. Therefore, since the family of CP distributions with Lebesgue intensity is compact, to prove Theorem 2.6, it suffices to show that the centering is a continuous operation. This is done in [16, Lemmas 3.5 and 3.6].

Together with convexity and the uniqueness of the ergodic decomposition, Theorem 2.6 implies that the family of fractal distributions is a Choquet simplex. Recall that a Poulsen simplex is a Choquet simplex in which extremal points are dense. Note that the set of extremal points is precisely the collection of ergodic fractal distributions. The following theorem is proved by Käenmäki, Sahlsten, and Shmerkin [16, Theorem B].

Theorem 2.7 *The family of fractal distributions is a Poulsen simplex.*

The proof is again based on the interplay between fractal distributions and CP processes. We prove that ergodic CP processes are dense by constructing a dense set of distributions of random self-similar measures on the dyadic grid. This is done by first approximating a given CP process by a finite convex combination of ergodic CP processes, and then, by splicing together those finite ergodic CP processes, constructing a sequence of ergodic CP processes converging to the convex combination. Roughly speaking, splicing of measures consists in pasting together a sequence of measures along dyadic scales. Splicing is often employed to construct measures with a given property based on properties of the component measures. For details, the reader is referred to [16, §4].

In geometric considerations, we usually construct a fractal distribution satisfying certain property. We often want to transfer that property back to a measure. This leads us to the concept of generated distributions.

Definition 2.8 (Uniformly scaling measures) We say that a measure μ *generates* a distribution P at x if

$$\mathcal{TD}(\mu, x) = \{P\}.$$

If μ generates P for μ-almost all x, then we say that μ is a *uniformly scaling measure*.

One can think that the uniformly scaling property is an ergodic-theoretical notion of self-similarity. Hochman proved the striking fact that generated distributions are always fractal distributions. The following result of Käenmäki, Sahlsten, and Shmerkin [16, Theorem C] is a converse to this.

Theorem 2.9 *If P is a fractal distribution, then there exists a uniformly scaling measure μ generating P.*

Recall that if P is an ergodic fractal distribution, then, by (2.1), P-almost every measure is uniformly scaling. Thus, by Theorems 2.6 and 2.7, it suffices to show that the collection of fractal distributions satisfying the claim is closed. Let $(P_i)_i$ be a sequence of ergodic fractal distributions converging to P and let μ_i be a uniformly scaling measure generating P_i. The proof is again based on the interplay between fractal distributions and CP processes. The rough idea to obtain a uniformly scaling measure generating P is to splice the measures μ_i together. For the full proof, the reader is referred to [16, §5].

3 Geometry of Measures

Let $G(d, d-k)$ denote the set of all $(d-k)$-dimensional linear subspaces of $\mathbb{R}^d$. For $x \in \mathbb{R}^d$, $r > 0$, $V \in G(d, d-k)$, and $0 < \alpha \leq 1$ define

$$X(x, r, V, \alpha) = \{y \in B(x, r) : \operatorname{dist}(y - x, V) < \alpha|y - x|\}.$$

Conical density results aim to give conditions on a measure which guarantee that the cones $X(x, r, V, \alpha)$ contain a large portion of the mass from the surrounding ball $B(x, r)$ for certain proportion of scales. For example, a lower bound on some dimension often is such a condition. Recall that the *lower local dimension* of a Radon measure μ at $x \in \mathbb{R}^d$ is

$$\underline{\dim}_{\text{loc}}(\mu, x) = \liminf_{r \downarrow 0} \frac{\log \mu(B(x, r))}{\log r} \tag{3.1}$$

and the *lower Hausdorff dimension* of μ is

$$\begin{aligned}\underline{\dim}_{\text{H}}(\mu) &= \operatorname*{ess\,inf}_{x \sim \mu} \underline{\dim}_{\text{loc}}(\mu, x) \\ &= \inf\{\dim_{\text{H}}(A) : A \subset \mathbb{R}^d \text{ is a Borel set with } \mu(A) > 0\}.\end{aligned}$$

Here $\dim_{\text{H}}(A)$ is the Hausdorff dimension of the set $A \subset \mathbb{R}^d$. A measure μ is *exact-dimensional* if the limit in (3.1) exists and is μ-almost everywhere constant. In this case, the common value is simply denoted by $\dim(\mu)$.

Intuitively, the local dimension of a measure should not be affected by the geometry of the measure on a density zero set of scales. Thus one could expect that tangent distributions should encode all information on dimensions.

Definition 3.1 (Dimension of fractal distributions) The *dimension of a fractal distribution P* is

$$\dim(P) = \int \dim(\mu) \, \mathrm{d}P(\mu).$$

The dimension above is well defined by the fact that if P is a fractal distribution, then P-almost every measure is exact-dimensional; see [12, Lemma 1.18]. The dimension of fractal distributions has also other convenient properties. While the Hausdorff dimension is highly discontinuous on measures, the function $P \mapsto \dim(P)$ defined on the family of fractal distributions is continuous; see [15, Lemma 3.20]. The usefulness of the definition is manifested in the following result of Hochman [12, Proposition 1.19]. Recall Theorem 2.5.

Theorem 3.2 *If μ is a Radon measure, then*

$$\underline{\dim}_{\text{loc}}(\mu, x) = \inf\{\dim(P) : P \in \mathcal{TD}(\mu, x)\}$$

for μ-almost all x. Furthermore, if μ is a uniformly scaling measure generating a fractal distribution P, then μ is exact-dimensional and $\dim(\mu) = \dim(P)$.

It turns out that tangent distributions are well suited to address problems concerning conical densities. The cones in question do not change under magnification and this allows to pass information between the original measure and its tangent distributions. Let

$$\mathcal{A}_\varepsilon := \{\nu \in \mathcal{M}_1 : \nu(X(0,1,V,\alpha)) \le \varepsilon \text{ for some } V \in G(d,d-k)\}$$

for all $\varepsilon \ge 0$. It is straightforward to see that $\mathcal{A}_\varepsilon$ is closed for all $\varepsilon \ge 0$; see [15, Lemma 4.2]. The key observation is that

$$\mathcal{A}_0 = \{\nu \in \mathcal{M}_1 : \mathrm{spt}(\nu) \cap X(0,1,V,\alpha) = \emptyset \text{ for some } V \in G(d,d-k)\},$$

where the defining property concerns only sets, is S_t invariant.

The following conical density result is proved by Käenmäki, Sahlsten, and Shmerkin [15, Proposition 4.3]. Roughly speaking, it claims that if the dimension of the measure is large, then there are many scales in which the cones contain a relatively large portion of the mass. A slightly more precise version is that there exists $\varepsilon > 0$ such that if $\underline{\dim}_{\mathrm{H}}(\mu) > k$, then for many scales $e^{-t} > 0$ we have

$$\inf_{V \in G(d,d-k)} \frac{\mu(X(x,e^{-t},V,\alpha))}{\mu(B(x,e^{-t}))} > \varepsilon$$

for μ-almost all x. The precise formulation of the theorem is as follows.

Theorem 3.3 *If* $k \in \{1,\ldots,d-1\}$, $k < s \le d$, *and* $0 < \alpha \le 1$, *then there exists* $\varepsilon > 0$ *satisfying the following: For every Radon measure* μ *on* $\mathbb{R}^d$ *with* $\underline{\dim}_{\mathrm{H}}(\mu) \ge s$ *it holds that*

$$\liminf_{T\to\infty} \langle\mu\rangle_{x,T}(\mathcal{M}_1 \setminus \mathcal{A}_\varepsilon) \ge \frac{s-k}{d-k}$$

for μ*-almost all* $x \in \mathbb{R}^d$.

The proof is based on showing that there cannot be "too many" rectifiable tangent measures. This means that, perhaps surprisingly, most of the known conical density results are, in some sense, a manifestation of rectifiability.

Definition 3.4 (Rectifiability) A set $E \subset \mathbb{R}^d$ is called *k-rectifiable* if there are countably many Lipschitz maps $f_i\colon \mathbb{R}^k \to \mathbb{R}^d$ so that

$$\mathcal{H}^k\Big(E \setminus \bigcup_i f_i(\mathbb{R}^k)\Big) = 0.$$

Here $\mathcal{H}^k$ is the k-dimensional Hausdorff measure. Observe that a k-rectifiable set E has $\dim_{\mathrm{H}}(E) \le k$. A sufficient condition for a set $E \subset \mathbb{R}^d$ to be k-rectifiable is that for every $x \in E$ there are $V \in G(d,d-k)$, $0 < \alpha < 1$, and $r > 0$ such that $E \cap X(x,r,V,\alpha) = \emptyset$; see [18, Lemma 15.13]. Thus, if a fractal distribution

P satisfies $P(\mathcal{A}_0) = 1$, then the quasi-Palm property implies that the support of P-almost every ν is k-rectifiable and hence $\dim(P) \le k$.

To prove Theorem 3.3, let $p, \delta > 0$ be such that $p < (s-\delta-k)/(d-k) < (s-k)/(d-k)$. Suppose to the contrary that there is $0 < \alpha \le 1$ so that for each $\varepsilon > 0$ there exists a Radon measure μ with $\underline{\dim}_{\mathrm{H}}(\mu) \ge s$ such that the claim fails to hold for p, that is,

$$\limsup_{T\to\infty} \langle \mu \rangle_{x,T}(\mathcal{A}_\varepsilon) > 1-p$$

on a set E_ε of positive μ measure. By Theorems 2.5 and 3.2, we may assume that at points $x \in E_\varepsilon$, all tangent distributions of μ are fractal distributions and

$$\inf\{\dim(P) : P \in \mathcal{TD}(\mu, x)\} = \underline{\dim}_{\mathrm{loc}}(\mu, x) > s-\delta.$$

Fix $x \in E_\varepsilon$. For each $\varepsilon > 0$, as $\mathcal{A}_\varepsilon$ is closed, we find a tangent distribution $P_\varepsilon \in \mathcal{TD}(\mu, x)$ so that $P_\varepsilon(\mathcal{A}_\varepsilon) \ge 1-p$. Since the sets $\mathcal{A}_\varepsilon$ are also nested, we get

$$P(\mathcal{A}_0) = \lim_{\varepsilon\downarrow 0} P(\mathcal{A}_\varepsilon) \ge 1-p,$$

where P is a weak limit of a sequence formed from P_ε as $\varepsilon \downarrow 0$. Furthermore, since the collection of all fractal distributions is closed by Theorem 2.6 and the dimension is continuous, the limit distribution P is a fractal distribution with

$$\dim(P) \ge s-\delta.$$

Let P_ω, $\omega \sim P$, be the ergodic components of P. By the invariance of $\mathcal{A}_0$, we have $P_\omega(\mathcal{A}_0) \in \{0,1\}$ for P-almost all ω. If $P_\omega(\mathcal{A}_0) = 0$, then we use the trivial estimate $\dim(P_\omega) \le d$, and if $P_\omega(\mathcal{A}_0) = 1$, then the rectifiability argument gives $\dim(P_\omega) \le k$. Since $P(\{\omega : P_\omega(\mathcal{A}_0) = 1\}) = P(\mathcal{A}_0) \ge 1-p$ we estimate

$$s-\delta \le \dim(P) = \int \dim(P_\omega)\,\mathrm{d}P(\omega) \le P(\mathcal{A}_0)k + (1-P(\mathcal{A}_0))d \le (1-p)k + pd$$

yielding $p \ge (s-\delta-k)/(d-k)$. But this contradicts the choice of δ. Thus the claim holds.

Relying on the existence of uniform scaling measures, we are able to study the sharpness of Theorem 3.3. The following result is proved by Käenmäki, Sahlsten, and Shmerkin [15, Proposition 4.4].

Theorem 3.5 *If* $k \in \{1, \ldots, d-1\}$*,* $k < s \le d$*, and* $0 < \alpha \le 1$*, then there exists a Radon measure* μ *on* $\mathbb{R}^d$ *with* $\dim(\mu) = s$ *such that*

$$\lim_{T\to\infty} \langle\mu\rangle_{x,T}(\mathcal{M}_1 \setminus \mathcal{A}_\varepsilon) = \begin{cases} (s-k)/(d-k), & \text{if } 0 < \varepsilon < \varepsilon(d,k,\alpha), \\ 0, & \text{if } \varepsilon > \varepsilon(d,k,\alpha), \end{cases}$$

for μ*-almost all* $x \in \mathbb{R}^d$*.*

Here, for $k \in \{1, \ldots, d-1\}$, $0 < \alpha \le 1$, and $V \in G(d, d-k)$, we have defined

$$\varepsilon(d,k,\alpha) := \frac{\mathcal{L}^d(X(0,1,V,\alpha))}{\mathcal{L}^d(B(0,1))}.$$

It follows from the rotational invariance of the Lebesgue measure $\mathcal{L}^d$ that $\varepsilon(d,k,\alpha)$ does not depend on the choice of V.

The measure μ above is just a uniform scaling measure generating

$$P = \frac{s-k}{d-k}\delta_{\mathcal{L}} + \Big(1 - \frac{s-k}{d-k}\Big)\delta_{\mathcal{H}},$$

where $\mathcal{L}$ is the normalization of $\mathcal{L}^d|_{B_1}$ and $\mathcal{H}$ is the normalization of $\mathcal{H}^k|_{W\cap B_1}$ for a fixed $W \in G(d,k)$. Since P is a convex combination of two fractal distributions, it is a fractal distribution. The existence of μ is guaranteed by Theorem 2.9. Recalling Theorem 3.2, we see that μ is exact-dimensional and

$$\dim(\mu) = \dim(P) = \frac{s-k}{d-k}d + \Big(1 - \frac{s-k}{d-k}\Big)k = s.$$

The goal is to verify that μ has the claimed properties.

Fix $0 < \varepsilon < \varepsilon(d,k,\alpha)$. Since $\mathcal{L}(X(0,1,V,\alpha)) = \varepsilon(d,k,\alpha) > \varepsilon$ for all $V \in G(d, d-k)$ and $\mathcal{H}(X(0,1,W^\perp,\alpha)) = 0$ we have $P(\mathcal{M}_1 \setminus \mathcal{A}_\varepsilon) = (s-k)/(d-k)$. Thus, by the weak convergence, it follows that

$$\lim_{T\to\infty} \langle\mu\rangle_{x,T}(\mathcal{M}_1 \setminus \mathcal{A}_\varepsilon) = \frac{s-k}{d-k}.$$

In the case $\varepsilon > \varepsilon(d,k,\alpha)$ we can reason similarly.

References

1. T. Bedford, A.M. Fisher, On the magnification of Cantor sets and their limit models. Monatsh. Math. **121**(1–2), 11–40 (1996)
2. T. Bedford, A.M. Fisher, Ratio geometry, rigidity and the scenery process for hyperbolic Cantor sets. Ergod. Theory Dyn. Syst. **17**(3), 531–564 (1997)

3. T. Bedford, A.M. Fisher, M. Urbański, The scenery flow for hyperbolic Julia sets. Proc. Lond. Math. Soc. (3) **85**(2), 467–492 (2002)
4. A.S. Besicovitch, On the fundamental geometrical properties of linearly measurable plane sets of points. Math. Ann. **98**(1), 422–464 (1928)
5. A.S. Besicovitch, On linear sets of points of fractional dimension. Math. Ann. **101**(1), 161–193 (1929)
6. A.S. Besicovitch, On the fundamental geometrical properties of linearly measurable plane sets of points II. Math. Ann. **115**, 296–329 (1938)
7. A. Ferguson, J. Fraser, T. Sahlsten, Scaling scenery of $(\times m, \times n)$ invariant measures. Adv. Math. **268**, 564–602 (2015)
8. A.M. Fisher, Small-scale structure via flows, in *Fractal Geometry and Stochastics III*. Volume 57 of Progress in Probability (Birkhäuser, Basel, 2004), pp. 59–78
9. H. Furstenberg, Intersections of Cantor sets and transversality of semigroups, in *Problems in Analysis (Sympos. Salomon Bochner, Princeton University, Princeton, 1969)* (Princeton University Press, Princeton, 1970), pp. 41–59
10. H. Furstenberg, Ergodic fractal measures and dimension conservation. Ergod. Theory Dyn. Syst. **28**(2), 405–422 (2008)
11. M. Gavish, Measures with uniform scaling scenery. Ergod. Theory Dyn. Syst. **31**(1), 33–48 (2011)
12. M. Hochman, Dynamics on fractals and fractal distributions (2010, preprint). Available at http://arxiv.org/abs/1008.3731
13. M. Hochman, P. Shmerkin, Equidistribution from fractal measures. Invent. Math. (Published Online)
14. A. Käenmäki, On upper conical density results, in *Recent Developments in Fractals and Related Fields*, ed. by J.J. Benedetto, J. Barral, S. Seuret. Applied and Numerical Harmonic Analysis (Birkhäuser, Boston, 2010), pp. 45–54
15. A. Käenmäki, T. Sahlsten, P. Shmerkin, Dynamics of the scenery flow and geometry of measures. Proc. Lond. Math. Soc. **110**(5), 1248–1280 (2015)
16. A. Käenmäki, T. Sahlsten, P. Shmerkin, Structure of distributions generated by the scenery flow. J. Lond. Math. Soc. **91**(2), 464–494 (2015)
17. J.M. Marstrand, Some fundamental geometrical properties of plane sets of fractional dimensions. Proc. Lond. Math. Soc. (3) **4**, 257–301 (1954)
18. P. Mattila, *Geometry of Sets and Measures in Euclidean Spaces: Fractals and Rectifiability* (Cambridge University Press, Cambridge, 1995)
19. P. Mörters, D. Preiss, Tangent measure distributions of fractal measures. Math. Ann. **312**(1), 53–93 (1998)
20. D. Preiss, Geometry of measures in $\mathbf{R}^n$: distribution, rectifiability, and densities. Ann. Math. (2) **125**(3), 537–643 (1987)
21. U. Zähle, Self-similar random measures and carrying dimension, in *Proceedings of the Conference: Topology and Measure, V, Binz, 1987*. Wissenschaftliche Beiträge der Ernst-Moritz-Arndt Universität Greifswald (DDR) (Greifswald, 1988), pp. 84–87

The Shape of Anisotropic Fractals: Scaling of Minkowski Functionals

Philipp Schönhöfer and Klaus Mecke

Abstract The shape of fractals can be characterized by intrinsic volumes, so-called Minkowski functionals, which share with the common d-dimensional volume of spatial structures the property of being additive. Here, we study the effects of anisotropy on the scaling behavior beyond the fractal dimension by applying tensorial functionals. It can be shown that Minkowski tensors of anisotropic pre-fractals scale with additional subdimensions. In addition, for anisotropic pre-fractals even scalar Minkowski functionals exhibit multiple edge subterms which merge for the isotropic case.

Keywords Integral geometry • Fractals • Anisotropy • Minkowski functionals • DLA

1 Introduction

The concept of fractal dimension is a standard method of characterizing complex structures and processes [14]. This dimension d_f determines the scaling behavior of the volume of a set embedded in d-dimensional Euclidean space [3]. Fractal geometries are applied to all kind of spatially structured objects studied in geography [13], economy [2], biology [7], medicine [9], and physics [15, 22]. In physics especially phase transitions are determined by fractal dimensions – also called critical exponents [21]. For example, networks of voids in a porous material such as a sandstone exhibit fractal behaviour at the critical percolation density ρ_c of grains. Since sandstones are made out of grains, there is a typical size l which can be used to model the pore space by a grid of voxels as shown in Fig. 1. The shape of the fractal pore space depends on the shape of the grains and can be characterized by a volume-boundary ratio. Another example is shown in Fig. 2 from [8]: long chain alkanes form two-dimensional domains at a solid/gas interface. Here, the characterization of the spatial structure plays an important role in describing and modelling the relevant

P. Schönhöfer • K. Mecke (✉)
Friedrich Alexander University Erlangen-Nürnberg, Staudtstraße 7, D-91058 Erlangen, Germany
e-mail: philipp.schoenhoefer@fau.de; klaus.mecke@physik.uni-erlangen.de

C. Bandt et al. (eds.), *Fractal Geometry and Stochastics V*, Progress in Probability 70, DOI 10.1007/978-3-319-18660-3_3

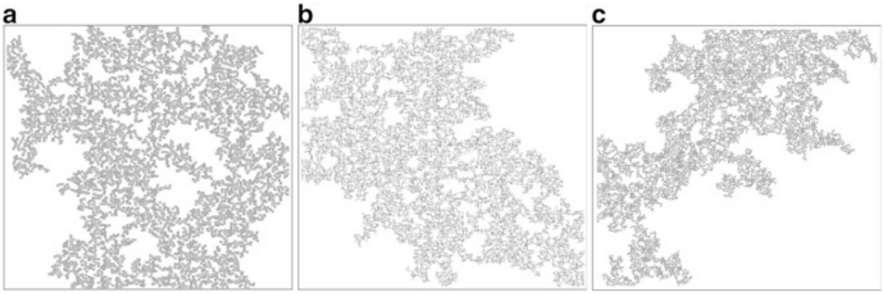


Fig. 1 Three percolating clusters of voids in a porous medium at the critical percolation threshold ρ_c. The typical size l of the grains determines the grid size. The shapes of the grains differ (**a**: *squares*, **b**: *sticks*, **c**: mixture of *squares* and *sticks*) which leads to different typical morphologies and anisotropies of the fractal percolating clusters (for more examples see [17])

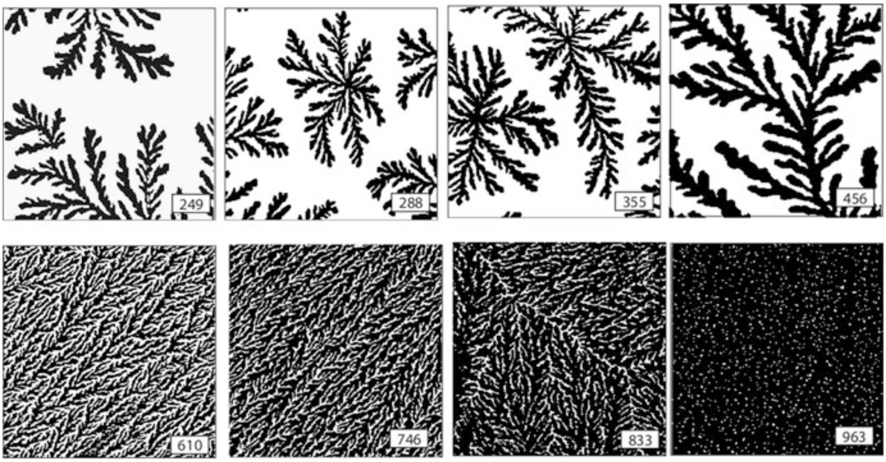


Fig. 2 Solid alkane domains (*black area*) at the surface of SiO_2 substrate (from [8]): the structures are fractal but the morphologies depend on the surface coverage (ranging between 24.9 % and 83.3 % in an area of typically 100 × 100 μm). The shape of the domains depends on the solidification conditions, i.e. the surface coverage, the cooling rate, etc. Thus, a morphological analysis may reveal the prevailing two-dimensional nucleation, transport, and solification processes. Typically, dendritic or seaweed shapes will appear due to morphological instabilities of the growth fronts which can be modelled by DLA-clusters shown in Fig. 3

physical and chemical processes. For instance, the morphology of the domains reflects the aggregation process of the particles. A typical and widespread example is the relation between diffusion-limited aggregation and fractal morphologies, which can be modelled by the Witten and Sander algorithm shown in Fig. 3.

The fractal structures studied in these applied research areas have one crucial issue in common: all objects in nature are built out from small elements of a certain size l which determines a smallest scale. Thus, from a mathematical point of view we deal with *pre-fractals* in these cases. It is, hence, "natural" to analyse the shape

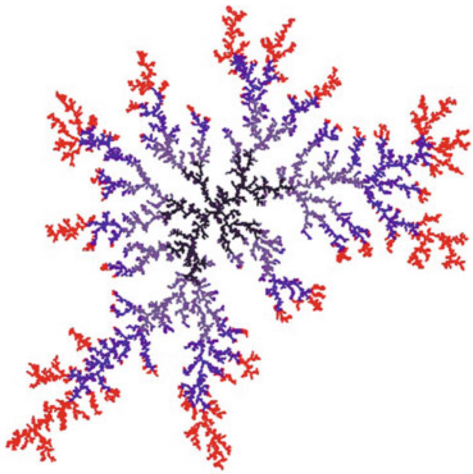


Fig. 3 The growth process of a DLA cluster which is modelled by the Witten and Sander algorithm after 1,000 (*black*), 3,300 (*grey*), 6,600 (*blue*), and 10,000 (*red*) particles of size l are added. The shape of the cluster, in particular, its anisotropy depends on the specific conditions of the growth process, e.g., the transport and solification of alkanes shown in Fig. 2

of fractals by exploiting a smallest scale l first and only afterwards study the limit $l \to 0$ – in contrast to the standard mathematical method [11, 18, 24], where the fractal limit $l \to 0$ is performed – before an analysis of its shape is done. Taking a finite value of l into account one can define scaling dimensions and amplitudes of *intrinsic volumes* or *Minkowski functionals*, which characterize the shape of fractals beyond the standard fractal dimension (see [16, 19]). Here, we study in particular the anisotropy of fractals by tensorial shape descriptors and restrict the analysis to two dimensions for the illustration of the technique.

1.1 Morphometry of Fractals

For most fractals the dimension d_f is indeed sufficient to characterize and thus to distinguish objects and processes. Nevertheless, sometimes more spatial information is needed to characterize shape. The Vicsek box fractals are two examples shown in Fig. 4 [4, 23]. Both fractals look similar: one is like the other but rotated by an angle of 45°. Thus in both cases the fractal dimension is $d_f = \frac{\ln 5}{\ln 3}$. However, the apparent shape is quite different because one is built out of crosses and one out of Xs. In order to distinguish these features the concept of a fractal dimension has to be extended. To differentiate the structures mathematically one may use intrinsic volumes and their scaling behavior, e.g. the boundary length V_1 and the Euler characteristics V_2 of a fractal $\mathcal{F}$ in two dimensions.

In [16] it was shown that the scaling of the intrinsic volumes V_1 and V_2 of a fractal $\mathcal{F}$, when observed within an observation windows W with different edge length x, exhibit *subterms* in addition to the fractal term visible in the 2-dimensional volume

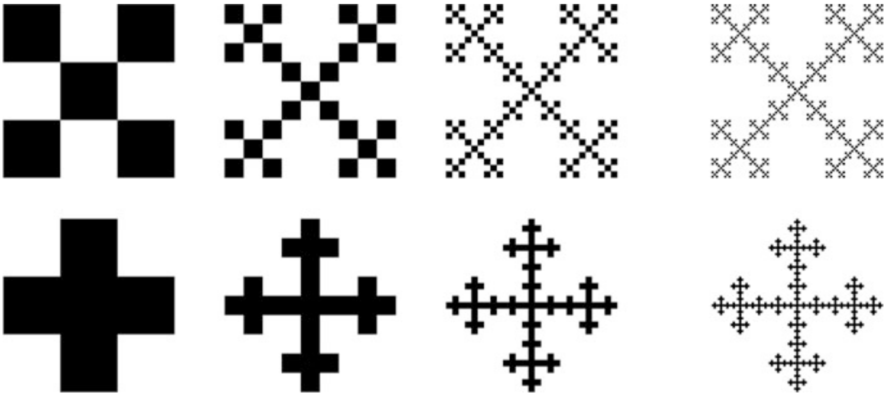


Fig. 4 Vicsek fractal constructed out of Xs (above) and out of crosses (below) in the zeroth, first, second, and fourth iterated step, from *left* to *right*

V_0, i.e.,

$$\begin{aligned} V_0(\mathcal{F} \cap W) &= v_{0;0}x^{d_f} \\ V_1(\mathcal{F} \cap W) &= v_{1;0}x^{d_f} + v_{1;1}x^{d_1} \\ V_2(\mathcal{F} \cap W) &= v_{2;0}x^{d_f} + v_{2;1}x^{d_1} + v_{2;2}x^{d_2}. \end{aligned} \tag{1.1}$$

The first exponent d_f is the conventional fractal dimension. The exponents d_1 and d_2 are called *subdimensions* and $v_{\nu;i}$ are the corresponding *amplitudes* of the dimension d_i. Here, we extend the analysis to anisotropic features of fractals by using tensorial Minkowski functionals.

1.2 Tensorial Minkowski Functionals

A spatial domain K with a smooth boundary contour ∂K can be characterized next to its scalar properties such as volume and boundary length also by tensorial additive valuations of its shape, the socalled tensorial Minkowski functionals or intrinsic volumes [5, 20]. In two-dimensional Euclidean space $\mathbb{E}^2$ the tensorial Minkowski functionals are defined by [20]

$$\begin{aligned} V_0^{a,b}(K) &= \int_K \vec{r}^a \otimes \vec{n}^b \mathrm{d}^2 r \\ V_1^{a,b}(K) &= \frac{1}{2}\int_{\partial K} \vec{r}^a \otimes \vec{n}^b \mathrm{d}r \\ V_2^{a,b}(K) &= \frac{1}{2}\int_{\partial K} \vec{r}^a \otimes \vec{n}^b \frac{\kappa}{\pi} \mathrm{d}r \, . \end{aligned} \tag{1.2}$$

where κ denotes the local curvature at the point $\vec{r} \in \partial K$ of the contour ∂K. For the normal vector $\vec{n}$ at the boundary point $\vec{r} \in \partial K$ the symmetric tensor products

$$\vec{r}^a \otimes \vec{n}^b = \underbrace{\vec{r} \otimes \cdots \otimes \vec{r}}_{a} \otimes \underbrace{\vec{n} \otimes \cdots \otimes \vec{n}}_{b}. \tag{1.3}$$

are used with $(\vec{r} \otimes \vec{n})_{ij} = \frac{r_i n_j + r_j n_i}{2}$. The functionals $V_\nu := V_\nu^{0,0}$ and $\vec{V}_\nu := V_\nu^{1,0}$ are called *Minkowski scalars* and *Minkowski vectors*, respectively. Thus, $V_0 := V_0^{0,0}$ is the volume of the body, $V_1 := V_1^{0,0}$ is proportional to the boundary length, and $V_2 := V_2^{0,0}$ describes the Euler characteristic. The Minkowski tensors are in general motion covariant, i.e., an important property of the tensorial Minkowski functionals is their behavior under translation [20]

$$V_\nu^{a,b}(K \uplus \vec{t}) = \sum_{i=0}^{a} \binom{a}{i} \vec{t}^i \otimes V_\nu^{a-i,b}(K) \tag{1.4}$$

where $K \uplus \vec{t}$ is a translation of K by $\vec{t}$. Alesker's theorem states [1], that the Minkowski tensors are complete set of additive tensorial functionals. However, not all possible rank-2 tensors has to be considered due to linear dependencies [5]

$$E_d V_\nu = \nu V_\nu^{0,2} + (n - \nu) V_{\nu+1}^{1,1} \tag{1.5}$$

where E_d denotes the rank-2 unit tensor in d dimensions. A list of the Minkowski rank-2 tensors used here is given by Table 1.

Table 1 A set of linear independent Minkowski tensors is given which describe shape and anisotropy of an object completely (see Eq. (1.5)) in regards to rank-2 tensors

Rank-2 Minkowski tensors 2D	Rank-2 Minkowski tensors 3D
V_0E_2	V_0E_2
V_1E_2	V_1E_2
V_2E_2	V_2E_2
–	V_3E_2
$V_1^{0,2}$	$V_1^{0,2}$
–	$V_2^{0,2}$
$V_0^{2,0}$	$V_0^{2,0}$
$V_1^{2,0}$	$V_1^{2,0}$
$V_2^{2,0}$	$V_2^{2,0}$
–	$V_3^{2,0}$

2 Minkowski Functionals of Fractals

Let us consider *iterated fractals* $\mathcal{F}$ which are constructed by an iterated algorithm [3, 6] with an *initiator* structure and an *iterated function system* (IFS) $\{\phi_1, \cdots, \phi_N\}$, such that the contracting maps ϕ_i obey

$$\bigcup_i \phi_i(I) \subseteq I \qquad \phi_i(I) \cap \phi_j(I) = \emptyset \text{ for } i \neq j. \tag{2.1}$$

The fractal set $\mathcal{F}$ is given after infinite steps

$$\bigcup_i \phi_i(\mathcal{F}) = \mathcal{F}. \tag{2.2}$$

Then, Minkowski functionals $V_\nu(\mathcal{F}, \epsilon)$ of the fractal set are calculated not directly of $\mathcal{F}$ but of the ϵ-neighborhood set [11, 18, 24]

$$F_\epsilon(\mathcal{F}) = \{x \in \mathbb{R}^d : \ \text{dist}(x, \mathcal{F}) < \epsilon\} \tag{2.3}$$

and by applying Steiner's formula [11]

$$V_0(F_\epsilon(\mathcal{F})) := \sum_{\nu=0}^{d} \binom{d}{\nu} V_\nu(\mathcal{F}, \epsilon)\epsilon^k. \tag{2.4}$$

In general, the Minkowski functionals $V_\nu(\mathcal{F}, \epsilon)$ oscillate with ϵ, but the limit $V_\nu(\mathcal{F}) = \lim_{\epsilon \to 0} V_\nu(\mathcal{F}, \epsilon)$ is well defined. However, the method has the disadvantages that it cannot be applied to the physical structures shown above and that one looses important information on the shape of the fractal structures found in Nature. Both is related to the existence of a smallest length scale l which regularises the fractal but does not exist for F_ϵ, where ϵ can be chosen arbitrarily small.

2.1 Regularisation by Smallest Length Scale

In physics fractal-like structures are constructed out of single particles which exhibit naturally a smallest length scale l. Therefore, we introduce an alternative method based on pre-fractals, which seems to be more 'natural' and allows for the definition of scaling amplitudes which characterize the shape of a fractal structure beyond the fractal dimension.

Let the initiator $I \subset \mathbb{R}^d$ be a closed hypercube around the origin $0^d = (\underbrace{0\,0 \cdots 0}_{d})^\top \in I$. For the first iterated step a set of N functions $\tilde{\phi}_i : \mathbb{R}^d \to \mathbb{R}^d$ is defined, which scales I by a real factor $s_i \geq 1$ and translates it by the vector $t_i \in \mathbb{R}^d$,

$$\begin{aligned} \tilde{\phi}_i(x) &= s_i \cdot x + t_i \qquad \forall i \in \{1, \ldots, N\} \\ \tilde{\phi}_i(I) \cap \phi_j(I) &= \emptyset \qquad \forall i, j \colon i \neq j. \end{aligned} \tag{2.5}$$

In addition we define a function $\Psi_r : \mathbb{R}^d \to \mathbb{R}^d$ with $\Psi_r(x) = r \cdot x$ and the *zoom factor* or *scaling factor*

$$r := \inf\{s \in \mathbb{R} \colon \bigcup_{i=1}^{N} \tilde{\phi}_i(I) \subseteq \Psi_s(I)\}. \tag{2.6}$$

Then, the functions of the n-th iterated step $\phi_i^{[n]}$ are given by

$$\phi_i^{[n]} \colon I^{[n-1]} \to I^{[n]} \qquad \phi_i^{[n]}(x) = s_i \cdot x + r^{n-1} t_i \tag{2.7}$$

where the sets $I^{[n]}$ are defined as

$$I^{[n]} := \Psi_r^{[n]} = \overbrace{\Psi_r \circ \cdots \circ \Psi_r}^{n}(I) \qquad I^{[0]}(I) = \mathrm{id}(I) = I. \tag{2.8}$$

For every iterated step the iterated function system is

$$\mathrm{IFS}^{[n]} := \{\phi_i^{[n]} \colon i \in \{1, \ldots, N\}\}. \tag{2.9}$$

The pre-fractal set F_n after n iterated steps is defined as the union $F_n := \bigcup_{i=1}^{N} \Phi_i^{[n]}(I)$ with

$$\Phi_i^{[n]}(I) = \phi_i^{[n]} \left(\bigcup_{i=1}^{N} \Phi_i^{[n-1]}(I) \right) \qquad \Phi_i^{[0]}(I) = \mathrm{id}(I). \tag{2.10}$$

The limit $F = F_\infty = \lim_{n\to\infty} F_n$ defines the fractal set F. Notice, that there is a regularization for every iterated step and even after infinite steps a smallest length scale is given by the size of the initiator I. This regularization ensures that the Minkowski functional $V_\nu(F)$ can be calculated without using neighborhood sets. The fractal dimension d_f is determined by the scaling of the volume V_0 with the iterated steps n. This procedure corresponds to the *Sandbox method* where F is investigated within different sized observation windows W of size $x = r^n$ with a scaling factor r [8]. The oscillations of $V_\nu(\mathcal{F}, \epsilon)$ in the previous method using Steiner's formula (2.4) are suppressed here. This alternative method can be viewed as an exchange of the limits $\epsilon \to 0$ and $n \to \infty$. The advantage of doing $\epsilon \to 0$ first is that shape descriptors such as the Minkowski functionals $V_\nu^{a,b}(n)$ exhibit a well-defined scaling behavior.

2.2 Isotropic Fractals

Minkowski functionals $V_\nu^{a,b}(F_n \cap W_x)$ of a pre-fractal F_n are calculated within a centered observation windows W_x with different lengths x [8]. For iterated fractals this corresponds to calculate the functionals for every iterated step $n = \log_r x$, where the *scaling factor* is symbolized by r. The initiator determines here the smallest scale size. The construction algorithm of the alternative, more 'natural' method is explained in detail in [19]. As in [16] the Minkowski functionals indeed reveal next to the fractal term additional *subterms* with *subdimensions* d_i and *intrinsic amplitudes* $v_{\nu;i}^{a,b}$. In general, the scalar functionals could be identified by the Eqs. (1.2) and the tensors of rank 2 by

$$\begin{aligned}
\mathrm{tr}[V_0^{2,0}](x) &= v_{0;0}^{2,0} x^{d_f+2} + v_{0;3}^{2,0} x^{d_3+2} \\
\mathrm{tr}[V_1^{2,0}](x) &= v_{1;0}^{2,0} x^{d_f+2} + v_{1;1}^{2,0} x^{d_1+2} + v_{1;3}^{2,0} x^{d_3+2} \\
\mathrm{tr}[V_2^{2,0}](x) &= v_{2;0}^{2,0} x^{d_f+2} + v_{2;1}^{2,0} x^{d_1+2} + v_{2;2}^{2,0} x^{d_2+2} + v_{2;3}^{2,0} x^{d_3+2} \\
\mathrm{tr}[V_1^{0,2}](x) &= v_{1;0}^{0,2} x^{d_f} + v_{1;1}^{0,2} x^{d_1} .
\end{aligned} \tag{2.11}$$

However, in [19] only fractals are considered which are isotropic, i.e., invariant under rotation of 90°. Here, we discuss the effects of anisotropy and analyse anisotropic iterated fractals, where rotation by an angle of 90° does not yield the same set. This is the case, for instance, in sandstones where the gravitational field breaks the orientational symmetry of the grains and thus of a percolating pore space (see Fig. 1), as well as in the process of diffusion limited growth (see Figs. 2 and 3), where patterned substrates or external fields may yield preferred orientations. The notation of [19] is used in the following.

3 Anisotropic Fractals: Minkowski Tensors of U-Fractals

Effects of anisotropy can be illustrated by the so-called U-fractal shown in Fig. 5. The initiator of this U-shaped fractal is the unit square $K_0 = [-0.5, 0.5]^2 \subset \mathbb{R}^2$. The initiator is mapped by seven functions $\phi_i(x) = s_i x + t_i$ with $s_i = 1$, where t_i are the seven different translation vectors with

$$t_i \in \left\{ \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}1\\-1\end{pmatrix}, \begin{pmatrix}0\\-1\end{pmatrix}, \begin{pmatrix}-1\\1\end{pmatrix}, \begin{pmatrix}-1\\0\end{pmatrix}, \begin{pmatrix}-1\\-1\end{pmatrix} \right\} . \tag{3.1}$$

Fig. 5 The U-fractal in the zeroth, first, second, and fourth iterated step, from *left* to *right*

The scaling factor of the carpet is $r = 3$. The elements of IFS$^{[n]}$ (see [19] Eq. (1.17)) for the n-th iterated step are

$$\phi_i^{[n]} : [-3^n \cdot 0.5, 3^n \cdot 0.5]^2 \to [-3^{n-1} \cdot 0.5, 3^{n-1} \cdot 0.5]^2 \qquad \phi_i^{[n]}(x) = x + 3^{n-1} t_i. \tag{3.2}$$

In the following, the Minkowski functionals $V_\nu^{(a,b)}$ are calculated for scalars $(a, b = 0)$, vectors $(a = 1,\ b = 0)$ and tensors of rank 2 $(a + b = 2)$. Compared to isotropic fractals we find additional subterms, which characterize the anisotopy of the fractal structure.

3.1 Minkowski Scalars

The Minkowski functionals are calculated analytically for every iterated step. First, we analyze the Minkowski scalars in dependence of the iterated step,

$$\begin{aligned} V_0(n) &= 7^n \\ V_1(n) &= \frac{7}{10}7^n + \frac{1}{2}3^n + \frac{4}{5}2^n \\ V_2(n) &= -\frac{2}{15}7^n + \frac{4}{5}2^n + \frac{1}{3}\,. \end{aligned} \tag{3.3}$$

The scaling behavior of the volume V_0 yields the fractal dimension $d_f = \log_3 7$. The other scalars, however, show deviations compared to the assumption of Eq. (1.2). The anisotropy reveals already one additional subterm here in the scalar functionals. This term can be identified as a second edge subterm which first occurs for V_1. The dimension of scaling terms can be identified by calculating the geometric or spectral zeta functions $\zeta_{\mathcal{F}}(s)$ and identify their singularities [10, 12]. Usually, the whole fractal structure $\mathcal{F}$ is considered to calculate $\zeta_{\mathcal{F}}(s)$. Here, we apply the technique also on subdimensional intersections, i.e., zeta functions of $F_n \cap E$ intersected with a hyperplane E, and define so called *edge zeta functions* $\zeta_E(s)$ of the pre-fractal F_n [19]. Then, it is immediately obvious why there are two subdimensions. For isotropic fractals like the Sierpiński carpet or the Vicsek fractals (see Fig. 4) the vertical and horizontal cutting lines where two maps $\phi_i^{[n]}(x)$ and $\phi_j^{[n]}(x)$ $i \neq j$ intersect in edges yield the same one-dimensional fractal string. In the case of the U-fractal two different fractal strings emerge (see Fig. 6). The first zeta function $\zeta_{E_1} = \zeta(s)$ corresponding to the vertical cut is the Riemann zeta function $\zeta(s)$, so that the first subdimension is the abscissa of convergence $d_1 = 1$ of $\zeta(s)$. The second zeta function ζ_{E_2} corresponding to the horizontal cut is

$$\zeta_{e_2}(s) = \frac{3^{-s}}{1 - 2 \cdot 3^{-s}}. \tag{3.4}$$

so that the second subdimension is consequently $d_2 = \log_3 2$. In the isotropic case these two dimensions merge. As expected the Euler characteristic V_2 has another subterm with $d_3 = 0$ corresponding to the intersection of four maps $\phi_i^{[n]}(x)$ in a vertex. Here, the isotropic and anisotropic fractal show no difference. Thus, the subdimensions coincide with the fractal dimensions of the structure in lower dimensional sections, i.e., of the pre-fractal edges and vertices, respectively.

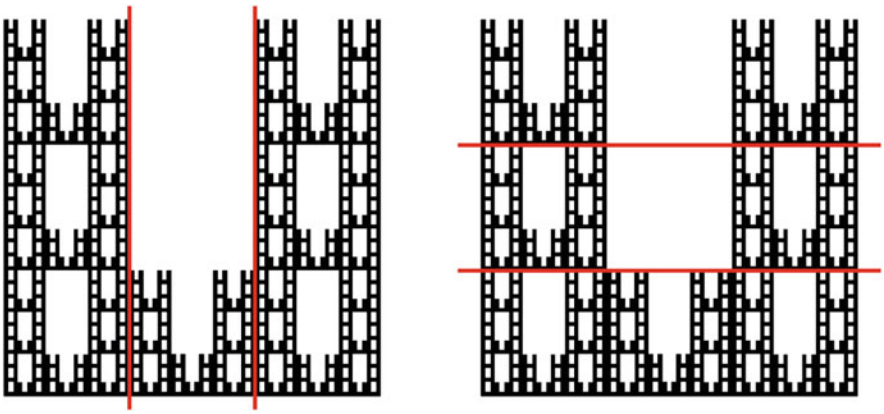


Fig. 6 U-fractal in the fourth iterated step. *Left*: *vertical cutting lines* are indicated; *right*: *horizontal cutting lines* are indicated

3.2 Minkowski Vectors

Due to anisotropy the Minkowski vectors $V_\nu^{1,0}$ are nontrivial and not negligible any more as in the isotropic case. The magnitudes of the vectors are

$$
\begin{aligned}
|V_0^{1,0}|(n) &= \frac{1}{14}21^n - \frac{1}{14}7^n \\
|V_1^{1,0}|(n) &= \frac{1}{20}21^n - \frac{3}{4}9^n + \frac{3}{2}7^n - \frac{4}{5}6^n \\
|V_2^{1,0}|(n) &= \frac{1}{105}21^n - \frac{11}{28}7^n + \frac{4}{5}6^n - \frac{5}{12}3^n \,.
\end{aligned}
\tag{3.5}
$$

The vectors scale as expected with the dimensions $d_0 + 1$, $d_1 + 1$, $d_2 + 1$, $d_3 + 1$, and d_0. The same effect was already calculated and explained for the rank-2 tensors in [19]. The summand $+1$ is naturally due to the position vector $\vec{r}$ in the integral Eq. (1.2) defining $V_\nu^{1,0}$ and should be subtracted from the scaling exponent in order to obtain the subdimensions. The vector $V_\nu^{1,0}$, consequently, scales with the subdimensions of V_ν and additionally with the vectorial dimension $d_4 = d_0 - 1$.

3.3 Minkowski Tensors

Finally, the tensors of rank 2 are calculated explicitly. The traces of the Minkowski tensors read

$$
\begin{aligned}
\mathrm{tr}[V_0^{2,0}](n) &= \frac{39}{196}63^n - \frac{1}{98}21^n - \frac{13}{588}7^n \\
\mathrm{tr}[V_1^{2,0}](n) &= \frac{39}{280}63^n + \frac{1}{15}27^n + \frac{3}{14}21^n + \frac{17}{110}18^n + \frac{1{,}157}{9{,}240}7^n - \frac{1}{30}2^n \\
\mathrm{tr}[V_2^{2,0}](n) &= -\frac{13}{490}63^n + \frac{11}{196}21^n + \frac{17}{110}18^n + \frac{3}{4}9^n - \frac{2{,}879}{5{,}390}7^n + \frac{1}{10}2^n \\
\mathrm{tr}[V_1^{0,2}](n) &= \frac{1}{20}21^n - \frac{3}{4}9^n + \frac{3}{2}7^n - \frac{4}{5}6^n.
\end{aligned}
\tag{3.6}
$$

Similar to the vectors the summand $+2$ has to be taken into account for the tensors $V_\nu^{2,0}$ because of the tensor product $\vec{r} \otimes \vec{r}$ in the integral in Eq. (1.2). Thus, the dimensions d_0, d_1, d_2, d_3 and $d_5 = d_0 - 2$ can be identified with the dimension which also occur in the isotropic case. In general, there are two additional tensorial subdimensions $d_6 = d_1 - 2$ and $d_7 = d_2 - 2$. The subdimension $d_2 - 2$ was already observed for the three dimensional Menger's sponge [19]. Since this subterm occurs only if $d_i \neq d - 1$ it cannot be observed for the Sierpiński carpet studied in [19]. The last dimension is the vectorial submission $d_4 = d_0 - 1$. Due to the covariant property

Table 2 Minkowski functionals of the U-fractal up to tensors of rank 2 (see Eq. (4.1–4.3)): the fractal dimension $d_f = d_0$, the subdimensions $d_{\nu\geq 1}$, and the corresponding amplitudes $v^{a,b}_{\mu;\nu}$ are obtained from Eqs. (3.3), (3.5), and (3.6), respectively

	Scalars				Vectors	Rank-2 tensors		
ν	0	1	2	3	4	5	6	7
d_ν	$\log_3 7$	1	$\log_3 2$	0	$d_0 - 1$	$d_0 - 2$	$d_1 - 2$	$d_2 - 2$
$v_{0;\nu}$	1	–	–	–	–	–	–	–
$v_{1;\nu}$	$\frac{7}{10}$	$\frac{1}{2}$	$\frac{4}{5}$	–	–	–	–	–
$v_{2;\nu}$	$-\frac{2}{15}$	0	$\frac{4}{5}$	$\frac{1}{3}$	–	–	–	–
$v^{1,0}_{0;\nu}$	$\frac{1}{14}$	–	–	–	$-\frac{1}{14}$	–	–	–
$v^{1,0}_{1;\nu}$	$\frac{1}{20}$	$-\frac{3}{4}$	$-\frac{4}{5}$	–	$\frac{3}{2}$	–	–	–
$v^{1,0}_{2;\nu}$	$\frac{1}{105}$	0	$\frac{4}{5}$	$-\frac{5}{12}$	$-\frac{11}{28}$	–	–	–
$v^{2,0}_{0;\nu}$	$\frac{39}{196}$	–	–	–	$-\frac{1}{98}$	$-\frac{13}{588}$	–	–
$v^{2,0}_{1;\nu}$	$\frac{39}{280}$	$\frac{3}{14}$	$\frac{17}{110}$	–	$\frac{1}{15}$	$\frac{1{,}157}{9{,}240}$	0	$-\frac{1}{30}$
$v^{2,0}_{2;\nu}$	$-\frac{13}{490}$	0	$\frac{17}{110}$	$\frac{3}{4}$	$\frac{11}{196}$	$-\frac{2{,}879}{5{,}390}$	0	$\frac{1}{10}$
$v^{0,2}_{1;\nu}$	$\frac{7}{10}$	$\frac{1}{2}$	$\frac{4}{5}$	–	–	–	–	–

(see Eq. (1.4)) the vectorial dimension also affects the tensors. Table 2 summarizes all subterms for the U-fractal, the scaling dimensions d_ν and the corresponding amplitudes $v^{a,b}_{\mu;\nu}$.

4 Conclusion

An alternative way of constructing and analyzing pre-fractals by Minkowski functionals was introduced in [19]. Here, we extended the method towards tensorial functionals and analyzed effects of anisotropy on the scaling behavior. To illustrate the method we calculated explicitly for an U-shaped iterated fractal the Minkowski functionals up to tensors of rank 2. In general, for anisotropic fractals the scaling behavior are for the Minkowski scalars

$$\begin{aligned} V_0(x) &= v_{0;0}x^{d_f} \\ V_1(x) &= v_{1;0}x^{d_f} + v_{1;1}x^{d_1} + v_{1;2}x^{d_2} \\ V_2(x) &= v_{2;0}x^{d_f} + v_{2;1}x^{d_1} + v_{2;2}x^{d_2} + v_{2;3}x^{d_3}, \end{aligned} \tag{4.1}$$

for the Minkowski vectors

$$\begin{aligned} |V_0^{1,0}|(x) &= v^{1,0}_{0;0}x^{d_f+1} + v^{1,0}_{0;4}x^{d_4+1} \\ |V_1^{1,0}|(x) &= v^{1,0}_{1;0}x^{d_f+1} + v^{1,0}_{1;1}x^{d_1+1} + v^{1,0}_{1;2}x^{d_2+1} + v^{1,0}_{1;4}x^{d_4+1} \\ |V_2^{1,0}|(x) &= v^{1,0}_{2;0}x^{d_f+1} + v^{1,0}_{2;1}x^{d_1+1} + v^{1,0}_{2;2}x^{d_2+1} + v^{1,0}_{2;3}x^{d_3+1} + v^{1,0}_{2;4}x^{d_4+1}, \end{aligned} \tag{4.2}$$

and for the Minkowski tensors of rank 2

$$\begin{aligned}
\mathrm{tr}[V_0^{2,0}](x) &= v_{0;0}^{2,0} x^{d_f+2} + v_{0;4}^{2,0} x^{d_4+2} + v_{0;5}^{2,0} x^{d_5+2} \\
\mathrm{tr}[V_1^{2,0}](x) &= v_{1;0}^{2,0} x^{d_f+2} + v_{1;1}^{2,0} x^{d_1+2} + v_{1;2}^{2,0} x^{d_2+2} + v_{1;4}^{2,0} x^{d_4+2} + v_{1;5}^{2,0} x^{d_5+2} \\
&\quad + v_{1;6}^{2,0} x^{d_6+2} + v_{1;7}^{2,0} x^{d_7+2} \\
\mathrm{tr}[V_2^{2,0}](x) &= v_{2;0}^{2,0} x^{d_f+2} + v_{2;1}^{2,0} x^{d_1+2} + v_{2;2}^{2,0} x^{d_2+2} + v_{2;3}^{2,0} x^{d_3+2} + v_{2;4}^{2,0} x^{d_4+2} \\
&\quad + v_{2;5}^{2,0} x^{d_5+2} + v_{2;6}^{2,0} x^{d_6+2} + v_{2;7}^{2,0} x^{d_7+2} \\
\mathrm{tr}[V_1^{0,2}](x) &= v_{1;0} x^{d_f} + v_{1;1} x^{d_1} + v_{1;2} x^{d_2}.
\end{aligned} \tag{4.3}$$

The specific values of the scaling dimensions d_ν and the corresponding amplitudes $v_{\mu;\nu}^{a,b}$ are given in Table 2 for the U-fractal shown in Fig. 5.

In particular, the Minkowski *scalars* V_ν show next to the leading fractal scaling two edge subterms and one vertex subterm, which are identical for isotropic fractals. It is interesting to notice that the anisotropic fractal exhibits no further *vertex* subterm. Consequently, we conclude that the number of different orientations of ν-dimensional boundary planes determine the maximal number of these subterms. For example for three dimensional fractals there are maximally three face subdimensions, three edge subdimensions, and one vertex subdimension. The triangular shape of the set of equations in Eq. (1.2) becomes less visible: the number of subterms admittedly increases with each equation, but the number does not increase by 1 for the next highest intrinsic volume (see Eq. (4.1)).

The Minkowski *vectors* $V_\nu^{1,0}$ are nontrivial for the anisotropic case. The vector $V_\nu^{1,0}$ scales with the subdimensions of the corresponding scalar V_ν and an vectorial subdimension $d_f - 1$. However, one has to add the summand $+1$ due to the definition of the vectors (see Eq. (1.2)). The same holds for the Minkowski *tensors* of rank 2. The scaling of the tensor $V_\nu^{2,0}$ is also determined by the scalar subdimensions and the vectorial subdimension $d_f - 1$. But additionally there are tensorial subdimensions, i.e., $d_f - 2$ based on the fractal dimension d_f, and $d_i - 2$ based on the edge subdimensions d_i.

It would be interesting to apply this morphometric analysis on physical fractals such as percolation clusters (Fig. 1) and diffusion limited aggregates shown in Figs. 2 and 3.

References

1. S. Alesker, Description of continuous isometry covariant valuations on convex sets. Geom. Dedicata **74**, 241–248 (1999)
2. C.J.G. Evertsz, Fractal geometry of financial time series. Fractals (3) **3**, 609–616 (1995)
3. K.J. Falconer, *Fractal Geometry: Mathematical Foundations and Applications*, 2nd edn. (Wiley, Chichester, 2003)

4. S.J. Fraser, Solvable random-decimation model of cluster scaling. Phys. Rev. A **38**, 953–961 (1988)
5. D. Hug, R. Schneider, R. Schuster, Integral geometry of tensor valuations. Adv. Appl. Math. **41**, 482–509 (2008)
6. J. Hutchinson, Fractals and self-similarity. Indiana Univ. Math. J. **30**, 713–747 (1981)
7. N.C. Kenkel, D.J. Walker, Fractals in the biological sciences. Coenoses **11**, 77–100 (1996)
8. L. Knüfing, H. Schollmeyer, H. Riegler, K. Mecke, Fractal analysis methods for solid alkane monolayer domains at SiO_2/air interfaces. Langmuir **21**, 992–1000 (2005)
9. G. Landini, J.W. Rippin, How important is tumour shape? Quantification of the epithelial-connective tissue interface in oral lesions using local connected fractal dimension analysis. J. Pathol. (2) **179**, 210–217 (1996)
10. M.L. Lapidus, M. van Frankenhuijsen, *Fractal Geometry, Complex Dimensions and Zeta Functions: Geometry and Spectra of Fractal Strings*, 1st edn. (Springer, New York, 2006)
11. M.L. Lapidus, E.P.J. Pearse, S. Winter, Pointwise tube formulas for fractal sprays and self-similar tilings with arbitrary generators. Adv. Math. **227**, 1349–1398 (2011)
12. M.L. Lapidus, C. Pomerance, Counterexamples to the modified Weyl-Berry conjecture on fractal drums. Math. Proc. Camb. Philos. Soc. **119**, 167–178 (1996)
13. B.B. Mandelbrot, How long is the coast of Britain? Statistical self-similarity and fractional dimension. Science **156**, 636–638 (1967)
14. B.B. Mandelbrot, *The Fractal Geometry of Nature*, 3rd edn. (W. H. Freeman and Company, New York, 1983)
15. C.J.A.P. Martins, E.P. Shellard, Fractal properties and small-scale structure of cosmic string networks. Phys. Rev. D **73**, 043515 (2006)
16. K. Mecke, *Additivity, Convexity, and Beyond: Applications of Minkowski Functionals in Statistical Physics.* Lecture Notes in Phys. vol. 554 (2000), pp. 111–184
17. K. Mecke, A. Seyfried, Strong dependence of percolation thresholds on polydispersity. Europhys. Lett. **58**, 28–34 (2002)
18. J. Rataj, S. Winter, Characterization of Minkowski measurability in terms of surface area. J. Math. Anal. Appl. **400**, 120–132 (2013)
19. P. Schönhöfer, K. Mecke, *Minkowski Functionals of Fractal Geometries* (2015, in preparation)
20. G.E. Schröder-Turk, S. Kapfer, B. Breidenbach, C. Beisbart, K. Mecke, Tensorial Minkowski functionals and anisotropy measures for planar patterns. J. Microsc. **238**, 57–74 (2010)
21. H.E. Stanley, *Introduction to Phase Transitions and Critical Phenomena*, 1st edn. (Oxford University Press, New York, 1971)
22. M. Suzuki, Phase transition and fractals. Prog. Theor. Phys. (1) **69**, 65–76 (1983)
23. T. Vicsek, Fractal models for diffusion controlled aggregation. J. Phys. A: Math. Gen. **16**, L647 (1983)
24. S. Winter, M. Zähle, Fractal curvature measures of self-similar sets. Adv. Geom. **13**, 229–244 (2013)

Projections of Self-Similar and Related Fractals: A Survey of Recent Developments

Pablo Shmerkin

Abstract In recent years there has been much interest -and progress- in understanding projections of many concrete fractals sets and measures. The general goal is to be able to go beyond general results such as Marstrand's Theorem, and quantify the size of *every* projection – or at least every projection outside some very small set. This article surveys some of these results and the techniques that were developed to obtain them, focusing on linear projections of planar self-similar sets and measures.

Keywords Self-similar sets • Self-similar measures • Projections • Hausdorff dimension • L^q dimensions

Mathematics Subject Classification (2000). Primary: 28A78, 28A80, Secondary: 37A99

1 Introduction

The study of the relationship between the Hausdorff dimension of a set and that of its linear projections has a long history, dating back to Marstrand's seminal projection theorem [24]:

Theorem 1.1 *Let $A \subset \mathbb{R}^2$ be a Borel set. Let Π_α denote the orthogonal projection onto a line making an angle α with the x-axis.*

(i) *If* $\dim_H A \leq 1$, *then* $\dim_H \Pi_\alpha A = \dim_H A$ *for almost every* $\alpha \in [0, \pi)$.
(ii) *If* $\dim_H A > 1$, *then* $\mathcal{L}(\Pi_\alpha A) > 0$ *for almost every* $\alpha \in [0, \pi)$.

The author was partially supported by Projects PICT 2011-0436 and PICT 2013-1393 (ANPCyT) http://www.utdt.edu/profesores/pshmerkin

P. Shmerkin (✉)
Department of Mathematics and Statistics, Torcuato Di Tella University, and CONICET, Buenos Aires, Argentina
e-mail: pshmerkin@utdt.edu

C. Bandt et al. (eds.), *Fractal Geometry and Stochastics V*, Progress in Probability 70, DOI 10.1007/978-3-319-18660-3_4

Here $\dim_{\mathrm{H}}$ stands for Hausdorff dimension and $\mathcal{L}$ for one-dimensional Lebesgue measure. Note that, in particular,

$$\dim_{\mathrm{H}}(\Pi_\alpha A) = \min(\dim_{\mathrm{H}} A, 1) \quad \text{for almost all } \alpha \in [0, \pi).$$

Although Marstrand's Theorem is very general, unfortunately it does not give information about what happens for a *specific* projection, and although the exceptional set is negligible in the sense of Lebesgue measure, it may still have large Hausdorff dimension. A more recent, and very active, line of research is concerned with gaining a better understanding of the size of projections of sets with some dynamical or arithmetic structure. The goal of this article is to present an overview of this area, focusing on projections of planar self-similar sets and measures (projections of other fractals are briefly discussed in Sect. 7). For a wider view of the many ramifications of Marstrand's Projection Theorem, the reader is referred to the excellent survey [7] in this volume.

2 Self-Similar Sets and Their Projections

We review some standard terminology and fix notation along the way. An **iterated function system** (IFS) on $\mathbb{R}^d$ is a finite collection $\mathcal{F} = (f_i)_{i\in\Lambda}$ of strictly contractive self-maps of $\mathbb{R}^d$. As is well known, for any such IFS $\mathcal{F}$, there exists a unique nonempty compact set A (the **attractor** or **invariant set** of $\mathcal{F}$) such that $A = \cup_{i\in\Lambda} f_i(A)$. We will repeatedly make use of the iterated IFS

$$\mathcal{F}^k = (f_i)_{i\in\Lambda^k}, \quad \text{where } f_{i_1\cdots i_k} = f_{i_1} \circ \cdots \circ f_{i_k},$$

which has the same attractor as $\mathcal{F}$. When the maps f_i are similarities, the set A is a **self-similar set**. From now on the maps f_i will always be assumed to be similarities, unless otherwise noted. For further background on self-similar sets and fractal dimensions, see e.g. [6].

Although we will be concerned with projections of self-similar sets, it will be useful to recall some ideas that apply to self-similar sets themselves. The **similarity dimension** $\dim_{\mathrm{S}}(\mathcal{F})$ of an IFS $\mathcal{F} = (f_i)_{i\in\Lambda}$ is the only positive root s of $\sum_{i\in\Lambda} \rho(f_i)^s = 1$, where $\rho(f)$ is the contraction ratio of the similarity f. If A is the attractor of $\mathcal{F}$, then there is a natural family of covers of A, namely

$$\{f_i(A) : i \in \Lambda^k\}.$$

Using these families, one can easily check that $\dim_{\mathrm{H}} A \le \dim_{\mathrm{S}}(A)$ (we follow a standard abuse of notation and speak of the similarity dimension of a self-similar set whenever the generating IFS is clear from context). Intuitively, it appears that if the sets $f_i(A), i \in \Lambda$ do not overlap much, then these covers should be close to optimal, and one should have an equality $\dim_{\mathrm{H}} A = \dim_{\mathrm{S}} A$. Recall that the IFS

$\mathcal{F}$ satisfies the **open set condition (OSC)** if there exists a nonempty open set O such that $f_i O \subset O$ for all i, and the images $f_i(O)$ are pairwise disjoint. The open set condition ensures that the overlap between the pieces $f_i(A)$ is negligible in a certain sense, and it is well known that Hausdorff and similarity dimensions agree whenever it holds. On the other hand, there are two trivial mechanisms that force the Hausdorff dimension to drop below the similarity dimension:

1. If $A \subset \mathbb{R}^d$ and $\dim_{\mathrm{S}}(A) > d$, then certainly $\dim_{\mathrm{H}}(A) \leq d < \dim_{\mathrm{S}}(A)$.
2. If $f_i = f_j$ for some $i \neq j$, then one can drop f_j from the IFS, resulting in a new generating IFS with strictly smaller similarity dimension. The same happens if two maps of $\mathcal{F}^k$ agree for some k, and in turn this happens if and only the semigroup generated by the f_i is not free. In this case we say that $\mathcal{F}$ has an **exact overlap**.

When the open set condition fails, but there are no exact overlaps, the combinatorial structure of the overlaps is very intricate, and calculating the dimension becomes much more challenging. In dimension $d = 1$, a major conjecture in the field is whether these are the *only* possible mechanisms for a drop in the Hausdorff dimension of a self-similar set (in higher dimensions this is false, but there is an analogous, albeit more complicated, conjecture).

We now turn our attention to *projections* of self-similar sets. Let A be a self-similar set generated by $(f_i)_{i\in\Lambda}$. If the similarities f_i are homotheties, i.e. $f_i(x) = \lambda_i x + t_i$ for some contractions $\lambda_i \in (0,1)$ and translations $t_i \in \mathbb{R}^d$, then for any linear map $\Pi : \mathbb{R}^d \to \mathbb{R}^k$, the image ΠA is also self-similar: it is the attractor of $(\lambda_i x + \Pi t_i)_{i\in\Lambda}$. We note that even if the original self-similar set satisfies the open set condition, their projections need not satisfy it; some of them (albeit only countably many) may have exact overlaps. In general, linear projections of self-similar sets need not be self-similar.

From now on we settle on the case $d = 2, k = 1$. We will say that a planar IFS $\mathcal{F}$ (or its attractor) is of **irrational type** if, for some k, $\mathcal{F}^k$ contains a map of the form $\lambda R_\theta x + t$ with θ/π irrational, where R_θ is rotation by θ. Otherwise, we say that the IFS is of **rational type**. We also say that $\mathcal{F}$ is **algebraic** if, when representing $f_i(x) = S_i x + t_i$ for a matrix $S_i \in \mathbb{R}^{2\times 2}$ and $t_i \in \mathbb{R}^2$, all the entries of S_i, t_i are algebraic for all $i \in \Lambda$. The following theorem summarizes the current knowledge about the projections of planar self-similar sets.

Theorem 2.1 *Let A be a planar self-similar set.*

(i) *If A is of irrational type, then* $\dim_{\mathrm{H}} \Pi_\alpha A = \min(\dim_{\mathrm{H}} A, 1)$ *for all* α.
(ii) *If A is algebraic, then* $\{\alpha : \dim_{\mathrm{H}} \Pi_\alpha A < \min(\dim_{\mathrm{H}} A, 1)\}$ *is countable.*
(iii) $\dim_{\mathrm{H}} \Pi_\alpha A = \min(\dim_{\mathrm{H}} A, 1)$ *for all* α *outside of a set of zero Hausdorff (and even packing) dimension.*
(iv) *If* $\dim_{\mathrm{H}} A > 1$, *then* $\dim_{\mathrm{H}}\{\alpha : \mathcal{L}(\pi_\alpha A) = 0\} = 0$.

Hence, without any assumptions, the exceptional set in Marstrand's Theorem for planar self-similar sets has zero Hausdorff dimension (rather than just zero Lebesgue measure).

Part (i) of Theorem 2.1 was proved by Peres and the author [29], with a different proof yielding many generalizations obtained later in [20]. We discuss a different approach in Sect. 4.

Claims (ii) and (iii) are consequences of some deep recent results of M. Hochman [18]. We present their proof, modulo a major result from [18], in Sect. 5.

The last part, concerning positive Lebesgue measure, was recently obtained by the author and B. Solomyak [33]. We will outline the proof in Sect. 6.

We will also discuss variants valid for (some) self-similar measures, which in most cases are a necessary step towards the proof of the set statements.

In the algebraic, rational type case, the set of exceptional directions can sometimes be explicitly determined. In particular, this is the case for the one-dimensional Sierpiński gasket, resolving a conjecture of Furstenberg. See Sect. 5.2 below.

We comment on the related natural question of what is the Hausdorff *measure* of $\Pi_\alpha A$ in its dimension. When $\dim_{\mathsf{H}} A > 1$, a partial answer is provided by Theorem 2.1(iv). When $\dim_{\mathsf{H}} A \leq 1$, in the irrational type case the answer is zero for all α. This was proved by Eroğlu [5] under the OSC (his result predates 2.1(i); he actually proved that the $\dim_{\mathsf{H}}(A)$-Hausdorff measure is zero), and recently extended to the general case by Farkas [11].

3 Dimension and Projection Theorems for Measures

3.1 Dimensions of Measures

Even if one is ultimately interested in sets, the most powerful methods for studying dimensions of projections involve measures in a natural way. Since a given set may support many dynamically relevant measures (such as self-similar or Gibbs measures) it is also useful to investigate measures for their own sake.

For sets, in this article we focus mostly on Hausdorff dimension. For measures, there are many notions of dimension which are useful or tractable, depending on the problem under consideration. We quickly review the ones we will need. From now on, by a measure we always mean a Radon measure (that is, locally finite and Borel regular) on some Euclidean space $\mathbb{R}^d$.

Given $x \in \operatorname{supp} \mu$, we define the **lower and upper local dimensions** of μ at x as

$$\underline{\dim}(\mu, x) = \liminf_{r \searrow 0} \frac{\log \mu B(x,r)}{\log r},$$

$$\overline{\dim}(\mu, x) = \limsup_{r \searrow 0} \frac{\log \mu B(x,r)}{\log r}.$$

If $\underline{\dim}(\mu, x) = \overline{\dim}(\mu, x)$, we write $\dim(\mu, x)$ for the common value and call it *the* local dimension at x. Local dimensions are functions; in order to obtain a global quantity, one may look at the μ-essential supremum or infimum of the local dimension. This yields four different notions of dimension, out of which the following two are most relevant for studying the dimension of projections:

$$\dim_* \mu = \sup\{s : \underline{\dim}(\mu, x) \geq s \quad \text{for } \mu\text{-almost all } x\}$$
$$\operatorname{Dim}^* \mu = \inf\{s : \overline{\dim}(\mu, x) \leq s \quad \text{for } \mu\text{-almost all } x\}.$$

Note that the supremum and infimum in question are attained. In the literature, $\dim_*$ and Dim^* are known as the *lower Hausdorff dimension* and *(upper) packing dimension* of a measure, respectively. The terminology stems from the following alternative characterization, which is closely related to the mass distribution principle:

$$\dim_* \mu = \inf\{\dim_{\mathsf{H}} A : \mu(A) > 0\},$$
$$\operatorname{Dim}^* \mu = \sup\{\dim_{\mathsf{P}} A : \mu(\mathbb{R}^d \setminus A) = 0\}.$$

Here $\dim_{\mathsf{P}}$ denotes packing dimension. The measure μ is called **exact dimensional** if $\dim_* \mu = \operatorname{Dim}^* \mu$ or, alternatively, if $\dim(\mu, x)$ exists and is μ-a.e. constant. Many dynamically defined measures are exact dimensional, but we note that, in general, a fixed projection of an exact dimensional measure needs not be exact dimensional.

A rather different notion of dimension (or rather, a one parameter family of dimensions) is related to the scaling law of the moments of the measure. Namely, given $q \geq 0, q \neq 1$, write

$$I_q(\mu, r) = \int \mu(B(x, r))^{q-1} \, d\mu(x), \tag{3.1}$$
$$D_q(\mu) = \liminf_{r \searrow 0} \frac{\log I_q(\mu, r)}{(q-1) \log r}.$$

The numbers D_q are known as the L^q dimensions of the measure, and are an essential ingredient of the multifractal formalism. The function $q \mapsto D_q\mu$ is always non-increasing, and

$$D_q(\mu) \leq \dim_* \mu \quad \text{for all } q > 1.$$

We refer to [10] for the proof of these facts, as well as further background on the different notions of dimension of a measure and their relationships. We finish by remarking that the value $q = 2$ is particularly significant, and $D_2(\mu)$ is also known as the **correlation dimension** of μ.

3.2 Dimensions of Self-Similar Measures

If μ is a measure on $\mathbb{R}^d$ and $g : \mathbb{R}^d \to \mathbb{R}^k$ is a map, we denote the push-forward of μ under g by $g\mu$, that is, $g\mu(B) = \mu(g^{-1}B)$ for all Borel sets B. If $\mathcal{F} = (f_i)_{i\in\Lambda}$ is an IFS and $p = (p_i)_{i\in\Lambda}$ is a probability vector, then there is a unique Borel probability measure $\mu = \mu(\mathcal{F}, p)$ such that

$$\mu = \sum_{i\in\Lambda} p_i f_i \mu.$$

The measure μ is called the **self-similar measure** associated to the IFS $\mathcal{F}$ and the weight p. For convenience we always assume that $p_i > 0$ for all i (otherwise one may pass to the IFS formed by the maps $(f_i : p_i > 0)$). In this case, the topological support of μ is the self-similar set associated to $\mathcal{F}$.

Self-similar measures are always exact dimensional; this is a rather deep fact which (at least in some special cases) can be traced back to ideas of Ledrappier and Furstenberg; see [12] for a detailed proof. As is the case for sets, dimensions of self-similar measures are well understood under the open set condition. In this case, one has

$$\dim \mu = \frac{\sum_{i\in\Lambda} p_i \log p_i}{\sum_{i\in\Lambda} p_i \log \rho(f_i)}. \tag{3.2}$$

This is an instance of the heuristic formula “dimension=entropy/Lyapunov exponent”, which often holds for measures invariant under some kind of conformal dynamics.

Regarding L^q dimensions, under the OSC it holds that $D_q\mu = \tau(q)/(q-1)$, where $\tau(q)$ is the only real solution to

$$\sum_{i\in\Lambda} p_i^q \rho(f_i)^{-\tau(q)} = 1.$$

In the special case where $p_i = \rho(f_i)^s$ (where $s = \dim_S(\mathcal{F})$), it can be easily checked that $\dim \mu = D_q\mu = s$ for all $q > 0$. These are called the **natural weights**.

Just as for sets, the formulae for $\dim \mu$ and $D_q\mu$ given above are expected to “typically” hold even in the presence of overlaps. For this reason, we call the right-hand side of (3.2) the **similarity dimension** of μ, and denote it $\dim_S \mu$.

3.3 Projection Theorems for Measures

Theorem 1.1 has an analog for various notions of dimension of a measure. The standard potential-theoretic proof of Marstrand’s Theorem (due to Kaufman) immediately yields a projection theorem for the correlation dimension. Projection theorems for other notions of dimension of a measure were obtained by Hu and Taylor [21, Theorem 6.1], and Hunt and Kaloshin [22, Theorem 1.1]; we remark that they are still fairly straightforward deductions from the proof of Theorem 1.1 as presented in i.e. [25, Chapter 9].

Theorem 3.1 *Let* dim *denote one of* $\dim_*$ *or* D_q *where* $q \in (1, 2]$*, and let* μ *be a measure on* $\mathbb{R}^2$*. The following holds for almost all* α*:*

1. *If* $\dim \mu \leq 1$*, then* $\dim \Pi_\alpha \mu = \dim \mu$
2. *If* $\dim \mu > 1$*, then* $\Pi_\alpha \mu$ *is absolutely continuous.*

The theorem fails for Dim^* and for D_q if $q \notin (1, 2]$, see [22]. When $D_q \mu > 1$, it can be shown that in fact $\Pi_\alpha \mu$ has an L^q density. There is an analogous result valid in higher dimensions.

4 The Irrational Case: Dimension of Projections

4.1 Projections of Some Self-Similar Measures

In this section we discuss the main ideas behind a proof of Theorem 2.1(i). The proof we sketch is based on ideas from [26], and is a particular case of more general results in [17].

For the time being we assume that $f_i(x) = \lambda R_\theta x + t_i$, $i \in \Lambda$, for some $\lambda \in (0, 1)$, $\theta \in [0, \pi)$ with $\theta/\pi \notin \mathbb{Q}$, and $t_i \in \mathbb{R}^2$ are translations. In this case, we say that the IFS $(f_i)_{i \in \Lambda}$ is **homogeneous**. In other words, in a homogeneous IFS, the linear parts are the same for all maps. Fix a probability vector $(p_i)_{i \in \Lambda}$, and let μ be the corresponding self-similar measure. The key to our proof of Theorem 2.1(i) is the following result.

Theorem 4.1 *If* μ *is as above, then for any* $q \in (1, 2]$ *and any* $\alpha \in [0, 2\pi)$*,*

$$D_q(\Pi_\alpha \mu) = \min(D_q \mu, 1).$$

We indicate the main steps in the proof. The first main ingredient is the inequality

$$I_q(\Pi_\alpha \mu, \lambda^{k+\ell}) \leq C_q \, I_q(\Pi_\alpha \mu, \lambda^k) \, I_q(\Pi_{\alpha + k\theta} \mu, \lambda^\ell), \tag{4.1}$$

valid for $q > 1$ for some constant $C_q > 0$. Recall (3.1). This is a consequence of the self-similarity of μ. A result of the same kind, for self-similar (and even self-conformal) measures rather than their projections, was obtained by Peres and Solomyak [30, Equation (3.2)], and the proof here is similar. The homogeneity of the IFS is key in deriving this inequality.

We can rewrite (4.1) as

$$\varphi_{k+\ell}(\alpha) \le \varphi_k(\alpha) + \varphi_\ell(T^k\alpha),$$

where T is the θ rotation on the circle (identified with $[0, 2\pi)$), and

$$\varphi_k(\alpha) = \log I_q(P_\alpha\mu, \lambda^k) + \log C_q.$$

In other words, φ_k is a subadditive cocycle over T, which is a uniquely ergodic transformation (this is where the irrationality of θ/π gets used). A result of Furman [16, Theorem 1] on subadditive cocycles over uniquely ergodic transformations implies that for *all* $\alpha \in [0, 2\pi)$ and almost all $\beta \in [0, 2\pi)$,

$$\liminf_{k\to\infty} \frac{\varphi_k(\alpha)}{k(q-1)\log\lambda} \ge \lim_{k\to\infty} \frac{\varphi_k(\beta)}{k(q-1)\log\lambda}.$$

The limit in the right-hand side exists and is a.e. constant from general considerations (the subadditive ergodic theorem), but in this case we know it equals $\min(D_q\mu, 1)$ by Theorem 3.1 (it is easy to see that, in the definition of D_q, one can take the limit along the sequence λ^k). This is the step of the proof that uses that $q \le 2$. It follows that $D_q(\Pi_\alpha\mu) \ge \min(D_q\mu, 1)$ for all α. The opposite inequality is trivial since D_q does not increase under Lipschitz maps and cannot exceed the dimension of the ambient space. This concludes the sketch of the proof of Theorem 4.1.

We point out that the analog of Theorem 4.1 holds for arbitrary self-similar measures (of irrational type) in the plane, at the price of replacing L^q dimension by Hausdorff dimension. This is a particular case of [20, Theorem 1.6]. The problem of whether Theorem 4.1 remains valid in this setting, for any values of q, remains open.

4.2 Conclusion of the Proof

We can now finish the proof of Theorem 2.1(i). If A is a self-similar set for a homogeneous IFS satisfying the open set condition, then we know that the self-similar measure μ with the natural weights satisfies $D_2\mu = \dim_S A = \dim_H A$, and hence, by Theorem 4.1,

$$\dim_H(\Pi_\alpha A) \ge D_2\Pi_\alpha\mu = \min(D_2\mu, 1) = \min(\dim_H(A), 1) \ge \dim_H(\Pi_\alpha A).$$

This shows that the claim holds when A has this special structure. To conclude the proof, we show that any self-similar set can be approximated in dimension from inside by such a self-similar set; this essentially goes back to [29]. We include the proof, since similar approximation arguments have turned to be useful in a variety of situations, see e.g. [27, Lemma 3.4], [11, Proposition 1.8], and [34, Section 9]. Recall that an IFS $(f_i)_{i\in\Lambda}$ with attractor A satisfies the **strong separation condition (SSC)** if the images $f_i(A)$ are pairwise disjoint (this is stronger than the OSC).

Lemma 4.2 *Let A be a self-similar set in $\mathbb{R}^2$ with $\dim_{\mathsf{H}} A > 0$. Then for any $\varepsilon > 0$ there is a self-similar set A' for a homogeneous IFS satisfying the strong separation condition, such that $A' \subset A$ and $\dim_{\mathsf{H}} A' \geq \dim_{\mathsf{H}} A - \varepsilon$.*

Proof Write $s = \dim_{\mathsf{H}} A$. It is classical that self-similar sets can be approximated in dimension from inside by self-similar sets satisfying the SSC. For completeness we sketch the argument: given $\varepsilon > 0$, we can find $r > 0$ arbitrarily small and a disjoint collection of balls $\{B(x_i, r)\}$ centres in A, with at least $r^{\varepsilon/2-s}$ elements. Since $x_i \in A$, it is easy to see that for each i there is a word j_i such that $f_{j_i}(A) \subset B(x_i, r)$ and $\rho(f_{j_i}) \geq \delta r$, where δ is a positive constant that depends only on the IFS. Then (f_{j_i}) satisfies the strong separation condition, the attractor A'' is contained in A, and its similarity dimension (equal to $\dim_{\mathsf{H}} A''$) can be made larger than $s - \varepsilon$ by taking r small enough. Hence, we may and do assume that A itself already satisfies the SSC.

A similarity f on $\mathbb{R}^2$ can be written as $\lambda O R_\theta + t$, where $\lambda \in (0,1)$, O is either the identity or reflection around the x-axis, R_θ is rotation by angle θ, and $t \in \mathbb{R}^2$ is a translation. Let Rot denote the similarities with O equal to the identity, and let Ref be the remaining ones. We claim that A can be approximated from inside by the attractor of an IFS with elements in Rot (that still satisfies the SSC). To see this, assume without loss of generality that $f_1 \in$ Ref. Fix a large integer k, and consider the IFS

$$\mathcal{F}'_k = (g \in \mathcal{F}^k \cap \mathrm{Rot}) \cup (f_1 g : g \in \mathcal{F}^k \cap \mathrm{Ref}).$$

A calculation shows that $\dim_{\mathsf{S}}(\mathcal{F}'_k)$ can be made arbitrarily close to s by taking k large enough, so this is the desired IFS.

Thus, we assume $\mathcal{F}$ satisfies the SSC and $f_i(x) = \lambda_i R_{\theta_i} x + t_i$ for suitable $\lambda_i \in (0,1)$, $\theta_i \in [0, 2\pi)$ and $t_i \in \mathbb{R}^2$. Because the similarities λR_θ commute, if we write $\mathcal{F}^k = (f_i(x) = S_i x + t_{k,i})_{i\in\Lambda^k}$, then S_i is determined by the number of times each index $\ell \in \Lambda$ appears in $i = (i_1, \ldots, i_k)$, whence there are fewer than $k^{|\Lambda|}$ different possibilities for S_i. Hence, there is some fixed similarity S, such that the IFS $\mathcal{F}'_k = (S_i x + t_{k,i} : S_i = S)$ satisfies $N\rho(S)^s \geq k^{-|\Lambda|}$, where N is the number of maps in $\mathcal{F}'_k$. On the other hand, $\rho(S) \leq \lambda_{\max}^k$, where $\lambda_{\max} = \max_{i\in\Lambda} \lambda_i < 1$. Hence, if we write $\dim_{\mathsf{S}}(\mathcal{F}'_k) = s - \varepsilon_k$, then

$$1 = N\rho(S)^{s-\varepsilon_k} \geq k^{-|\Lambda|}\rho(S)^{-\varepsilon_k} \geq k^{-|\Lambda|}\lambda_{\max}^{-\varepsilon_k k}.$$

Thus $(1/\lambda_{\max})^{\varepsilon_k k} \leq k^{|\Lambda|}$, and therefore $\varepsilon_k \to 0$ as $k \to \infty$. Since $\mathcal{F}'_k$ is a homogeneous IFS satisfying the SSC, whose attractor is contained in A (as it is derived from $\mathcal{F}^k$ by deleting some maps), this completes the proof. □

5 The Rational Rotation Case: Hochman's Theorem on Super Exponential Concentration

5.1 Hochman's Theorem

In the introduction we briefly discussed the following conjecture:

Conjecture 5.1 If A is a self-similar set in $\mathbb{R}$ with $\dim_{\mathsf{H}}(A) < \min(\dim_{\mathsf{S}}(A), 1)$, then A has exact overlaps.

Although a full solution to the conjecture seems to be beyond reach of current methods, a major breakthrough was recently achieved by M. Hochman [18]. Hochman proved a weaker form of the conjecture, which allows to establish the full conjecture in a number of important special cases. In order to state his result, we need some definitions. The **separation constant** of an IFS $\mathcal{F} = (\lambda_i x + t_i)_{i\in\Lambda}$ on the real line is defined as

$$\Delta(\mathcal{F}) = \begin{cases} \infty & \text{if } \lambda_i \neq \lambda_j \text{ for all } i \neq j, \\ \min_{i\neq j}\{|t_i - t_j| : \lambda_i = \lambda_j\} & \text{otherwise} \end{cases}.$$

Although $\Delta(\mathcal{F})$ may be infinite, we will only be interested in $\Delta(\mathcal{F}^k)$ for large values of k, and this is always finite (already for $k = 2$) due to the commutativity of the contraction ratios. The sequence $k \mapsto \Delta(\mathcal{F}^k)$ is always decreasing. Notice also that there is an exact overlap if and only if $\Delta(\mathcal{F}^k) = 0$ for some (and hence all sufficiently large) k. On the other hand, by pigeonholing it is easy to see that $\Delta(\mathcal{F}^k)$ decays at least exponentially fast in k. We say that $\mathcal{F}$ has **superexponential concentration of cylinders (SCC)** if $\Delta(\mathcal{F}^k)$ decays at superexponential speed or, in other words, if

$$\lim_{k\to\infty} \frac{-\log \Delta(\mathcal{F}^k)}{k} = \infty.$$

We can now state Hochman's Theorem:

Theorem 5.2 *If $\mu = \mu(\mathcal{F}, p)$ is a self-similar measure in $\mathbb{R}$ with $\dim_{\mathsf{H}} \mu < \min(\dim_{\mathsf{S}} \mu, 1)$, then $\mathcal{F}$ has super-exponential concentration of cylinders.*

In particular, if A is a self-similar set in $\mathbb{R}$ with $\dim_{\mathsf{H}}(A) < \min(\dim_{\mathsf{S}}(A), 1)$, then (the IFS generating) A has super-exponential concentration of cylinders.

The proof of this result combines several major new ideas. A key ingredient is an inverse theorem for the growth of entropy under convolutions, which belongs to the field of additive combinatorics. See the survey [19] or the introduction of [18] for an exposition of the main ideas in the proof.

In the remainder of this section, we explain how to apply Theorem 5.2 to the calculation of the dimension of projections of planar self-similar sets and measures, in the rational rotation case. In turn, this will be a key ingredient for establishing absolute continuity of projections. For many other applications of Theorem 5.2, see [15, 28, 32, 33] in addition to [18].

5.2 Projections of the One-Dimensional Sierpiński Gasket, and Theorem 2.1(ii)

The attractor S of the planar IFS $\left((\frac{x}{3}, \frac{y}{3}), (\frac{x+1}{3}, \frac{y}{3}), (\frac{x}{3}, \frac{y+1}{3})\right)$ is known as the **one-dimensional Sierpiński Gasket**, see Fig. 1. Since S satisfies the SSC, indeed $\dim_{\mathsf{H}}(S) = 1$. Because the generating IFS has no rotations, the orthogonal projections of S onto lines are again self-similar sets. Let $P_u(x, y) = x + uy$. Then $S_u := P_u S$ is homothetic to $\Pi_{\tan^{-1} u} S$. This provides a smooth reparametrization of the orthogonal projections of S, which has the advantage that S_u is the attractor of the simpler IFS $\left(\frac{x}{3}, \frac{x+1}{3}, \frac{x+u}{3}\right)$.

It is clear that S_u has an exact overlap for some values of u, for example, for $u = 1$. Kenyon showed that there is an exact overlap if and only if $u = p/q$ in lowest terms with $p + q \not\equiv 0 \bmod 3$, see [23, Lemma 6], and provided an expression

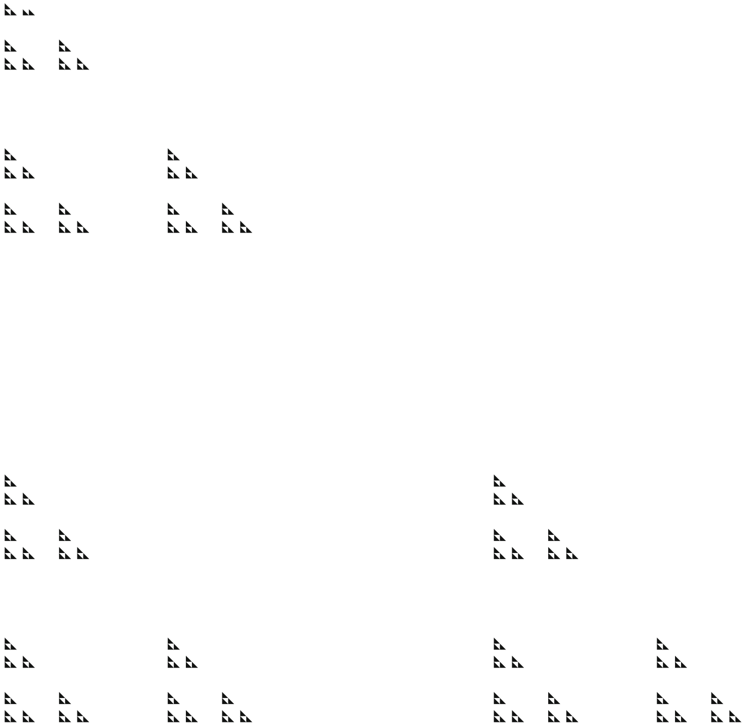


Fig. 1 The one-dimensional Sierpiński gasket

for the dimension of S_u in this case. Hence, unlike the irrational rotation case, there are exceptional directions, and an infinite number of them. Kenyon rounded off the understanding of projections with rational slope by showing that if $u = p/q$ in lowest terms with $p + q \equiv 0 \bmod 3$, then S_u has positive Lebesgue measure (in particular, dimension 1).

An old (unpublished) conjecture of H. Furstenberg states that $\dim S_u = 1$ for all *irrational* u. Since S_u has no exact overlap for u irrational, this is a particular case of Conjecture 5.1. In the same article [23], Kenyon proved that S_u has Lebesgue measure zero for all irrational u (answering a question of Odlyzko), and exhibited a dense G_δ set of irrational u such that $\dim_{\mathsf{H}} S_u = 1$. It turns out that a positive solution to Furstenberg's conjecture follows rather easily from Theorem 5.2 (the short deduction is due to Solomyak and the author). The argument is presented in [18, Theorem 1.6] just for projections of the one-dimensional Sierpiński gasket. A variant of the proof yields the following more general result.

Theorem 5.3 *Let $\lambda \in (0,1)$ be algebraic, and let a_i, b_i $(i \in \Lambda)$ also be algebraic. Suppose that the IFS*

$$\mathcal{F} = \big(\lambda(x + a_i), \lambda(y + b_i)\big)_{i\in\Lambda}$$

does not have an exact overlap. Write $s = \dim_{\mathsf{S}}(\mathcal{F}) = \log|\Lambda|/\log(1/\lambda)$. Let S be the attractor of $\mathcal{F}$, and let S_u be the image of S under $(x, y) \mapsto x + uy$.

Then $\dim_{\mathsf{H}} S_u = \min(s, 1)$ for all u such that the S_u does not have an exact overlap, and in particular for all but countably many u.

In the proof we will need the following lemma, see [18, Lemma 5.10] for the proof. Given a finite set B, the family of polynomial expressions in elements of B of degree at most k will be denoted $\mathcal{P}_k(B)$.

Lemma 5.4 *Let B be a finite set of algebraic numbers. There is a constant $\delta = \delta(B) > 0$ such that if $x \in \mathcal{P}_k(B)$, then either $x = 0$ or $|x| \geq \delta^k$.*

Proof of Theorem 5.3 The projection S_u is the attractor of

$$\mathcal{F}_u = \big(\lambda(x + a_i + ub_i)\big)_{i\in\Lambda}.$$

If u is algebraic, then $\mathcal{F}_u$ is algebraic (i.e. all the parameters are algebraic). As shown in [18, Theorem 1.5], it follows easily from Theorem 5.2 and Lemma 5.4 that in this case $\dim_{\mathsf{H}} S_u = \min(\dim_{\mathsf{S}} S_u, 1) \geq \min(s, 1)$ if $\mathcal{F}_u$ does not have an exact overlap. Hence from now on we assume that u is transcendental.

The IFS $\mathcal{F}_u^k$ is given by $(\lambda^k x + \sum_{\ell=1}^k \lambda^\ell(a_{i_\ell} + ub_{i_\ell}))_{i\in\Lambda^k}$. In particular, there are $i \neq j \in \Lambda^k$ such that

$$\Delta(\mathcal{F}_u^k) = x_k + uy_k,$$

where $x_k = \sum_{\ell=1}^k \lambda^\ell(a_{i_\ell} - a_{j_\ell})$, $y_k = \sum_{\ell=1}^k \lambda^\ell(b_{i_\ell} - b_{j_\ell})$. Since u is transcendental, this can be zero only if x_k and y_k are both zero, but in this case $\mathcal{F}^k$ can be seen to have an exact overlap, which contradicts our hypothesis. Hence for each k either $x_k \neq 0$ or $y_k \neq 0$.

Note that $x_k, y_k \in \mathcal{P}_{k+1}(B)$, where $B = \{\lambda, a_i - a_j, b_i - b_j : i,j \in \Lambda\}$. Let $\delta = \delta(B)$ be the number given by Lemma 5.4. If $x_k = 0$ or $y_k = 0$, then $|\Delta(\mathcal{F}_u^k)| \geq \min(1, |u|)\delta^{k+1}$. If this happens for infinitely many k then, in light of Theorem 5.2, $\dim_{\mathsf{H}} S_u = \min(\dim_{\mathsf{H}} S, 1)$ and we are done. Hence we may assume that $x_k y_k \neq 0$ for all $k \geq k_0$.

For any $k \geq k_0$, we hence have

$$\left| \frac{\Delta(\mathcal{F}_u^k)}{y_k} - \frac{\Delta(\mathcal{F}_u^{k+1})}{y_{k+1}} \right| = \left| \frac{x_k}{y_k} - \frac{x_{k+1}}{y_{k+1}} \right| = \left| \frac{z_k}{y_k y_{k+1}} \right|,$$

where $z_k \in \mathcal{P}_{2k+3}(B)$. Therefore, Lemma 5.4 yields that either $z_k = 0$ or $|z_k| \geq \delta^{2k+3}$. Assume first that $z_k = 0$ for all sufficiently large k, say for all $k \geq k_1 \geq k_0$. Then

$$|\Delta(F_u^k)| = |y_k(x_{k_1}/y_{k_1} + u)| \geq |x_{k_1}/y_{k_1} + u|\delta^{k+1} \quad \text{for all } k \geq k_1,$$

so there is no SCC and the conclusion follows again from Theorem 5.2. Thus, we may and do assume that $z_k \neq 0$ for infinitely many k. For any such k, since $|y_k|$ is bounded uniformly in k, we conclude that either

$$|\Delta(\mathcal{F}_u^k)| \geq c\delta^{2k+3}, \quad \text{or} \quad |\Delta(\mathcal{F}_u^{k+1})| \geq c\delta^{2k+3},$$

for some $c > 0$ independent of k. This shows that also in this case there is no SCC, so a final application of Theorem 5.2 finishes the proof. □

The same proof works for self-similar measures for $\mathcal{F}_u$.

Part (ii) of Theorem 2.1 follows from Theorem 5.3 and Lemma 4.2: if A is of irrational type there is nothing to do by part (i). If A is algebraic and of rational type then, given $\varepsilon > 0$, Lemma 4.2 provides us with an IFS $\mathcal{F}$ satisfying the hypotheses of Proposition Theorem 5.3 such that $\dim_{\mathsf{S}} \mathcal{F} > \dim_{\mathsf{H}} A - \varepsilon$ (it is clear from the proof that $\mathcal{F}$ is still algebraic, and also has rational rotations, so after iterating we may assume it has no rotations). To finish the proof we apply Theorem 5.3 to $\mathcal{F}$ and let $\varepsilon \searrow 0$ along a sequence.

5.3 Dimension of Projections in the Rational Case

We now apply Theorem 5.2 to prove Theorem 2.1(iii). Since we already know that in the irrational rotation case there are no exceptional directions at all, it remains to deal with the rational rotation case. Once again, we will first establish a corresponding result for measures, but imposing some additional structure on the IFS. Then we will deduce the general case for sets from Lemma 4.2.

Proposition 5.5 *Let $\mathcal{F} = (f_i)_{i\in\Lambda}$ be a planar IFS satisfying the SSC, where $f_i(x) = \lambda x + t_i$ for all i, that is, the maps f_i are homotheties with the same contraction ratio. There exists a set $E \subset [0,\pi)$ of zero Hausdorff (and even packing) dimension, such that if $\mu_p = \mu(\mathcal{F}, p)$ denotes the self-similar measure for $\mathcal{F}$ and the weight p, then*

$$\dim_{\mathsf{H}} \Pi_\alpha \mu_p = \min(\dim_{\mathsf{S}} \mu_p, 1) \quad \textit{for all } \alpha \in [0,\pi) \setminus E.$$

In particular, if A is the attractor of $\mathcal{F}$, then

$$\dim_{\mathsf{H}} \Pi_\alpha A = \min(\dim_{\mathsf{H}} A, 1) \quad \textit{for all } \alpha \in [0,\pi) \setminus E.$$

Proof The latter claim follows by applying the first claim to the natural weights. Once again, instead of working directly with orthogonal projections Π_α, we work with the family $P_u(x,y) = x + uy$; this makes no difference in the statement since the reparametrization is smooth and hence preserves Hausdorff and packing dimension.

As before, write $\mathcal{F}_u$ for the projected IFS $(\lambda x + P_u t_i)_{i\in\Lambda}$. Note that

$$\Delta(\mathcal{F}_u^k) = \min_{j\neq j'\in\Lambda^k} \Gamma_{j,j'}(u),$$

where

$$\Gamma_{j,j'}(u) = \left| \sum_{i=0}^{k-1} \lambda^i P_u t_{j_i} - \sum_{i=0}^{k-1} \lambda^i P_u t_{j'_i} \right| = \left| P_u \left(\sum_{i=0}^{k-1} \lambda^i (t_{j_i} - t_{j'_i}) \right) \right| =: |P_u(t_{j,j'})|.$$

Since $\mathcal{F}$ satisfies the SSC, $|t_{j,j'}| > c > 0$ for some $c = c(\mathcal{F})$. Let E denote the set of u such that $\mathcal{F}_u$ has SCC. Then, by definition of SCC,

$$E \subset \bigcap_{\varepsilon>0} \bigcup_{K=1}^{\infty} \bigcap_{k=K}^{\infty} \bigcup_{j\neq j'\in\Lambda^k} \Gamma_{j,j'}^{-1}(-\varepsilon^k, \varepsilon^k) =: \bigcap_{\varepsilon>0} E_\varepsilon.$$

Fix an interval $I = [-M, M]$. We first note that, since $|t_{j,j'}| > c$, for large enough k, the set $I \cap \Gamma_{j,j'}^{-1}(-\varepsilon^k, \varepsilon^k)$ can be covered by an interval of length $O(\varepsilon^k)$. Hence $\bigcup_{j\neq j'\in\Lambda^k} \Gamma_{j,j'}^{-1}(-\varepsilon^k, \varepsilon^k)$ can be covered by $|\Lambda|^{2k}$ intervals of length $O(\varepsilon^k)$, which implies that

$$\overline{\dim}_{\mathsf{B}} \left(\bigcap_{k=K}^{\infty} \bigcup_{j\neq j'\in\Lambda^k} \Gamma_{j,j'}^{-1}(-\varepsilon^k, \varepsilon^k) \right) \leq O\left(|\log\varepsilon|^{-1}\right),$$

with the implicit constant depending on $|\Lambda|$, where $\overline{\dim}_{\mathsf{B}}$ is upper box-counting (or Minkowski) dimension. In turn, since packing dimension is σ-stable and bounded above by $\overline{\dim}_{\mathsf{B}}$, this shows that $\dim_{\mathsf{P}}(E_\varepsilon \cap I) = O\left(|\log\varepsilon|^{-1}\right)$. Since $I = [-M, M]$

was arbitrary, we conclude that $\dim_{\mathsf{P}}(E) = 0$. The claim now follows from Theorem 5.2. □

The above proposition is a particular case of [18, Theorem 1.8], which deals with much more general analytic families of self-similar measures. The proof of the more general result is similar, except that in order to show that $\Gamma_{j,j'}^{-1}(-\varepsilon^k, \varepsilon^k)$ can be covered efficiently one needs to rely on "higher-order transversality" estimates (which are trivial in our setting because $\Gamma_{j,j'}$ is affine).

Claim (iii) of Theorem 2.1 follows from Proposition 5.5 and Lemma 4.2 in exactly the same way as part (ii) followed from Theorem 5.3.

6 Absolute Continuity of Projections

The methods from [18, 20, 26, 29] that we have discussed so far appear to be intrinsically about dimension (of sets or measures) and so far have not yielded any new information about positive Lebesgue measure or absolute continuity when the similarity dimension exceeds the dimension of the ambient space. Recently, in [32, 33] these results on dimension have been combined with some new ideas to yield absolute continuity outside a small set of parameters for many parametrized families of self-similar (and related) measures. One particular application of these ideas is the last claim of Theorem 2.1. In this section we discuss the main steps in the proof, referring the reader to [33] for the details.

We start by describing the general scheme for proving absolute continuity outside a small set of parameters. The measures μ_u to which the method applies have an infinite convolution structure: they are the distribution of a random sum

$$\sum_{n=1}^{\infty} X_{u,n},$$

where $X_{u,n}$ are independent Bernoulli random variables, and $\|X_{u,n}\|_\infty$ decreases exponentially uniformly in u, so that the series converges absolutely. Once a large integer k is fixed, this allows as to express μ_u as a convolution $\eta_u * \nu_u$, where η_u is the distribution of $\sum_{k|n} X_{u,n}$, and ν_u is the distribution of $\sum_{k\nmid n} X_{u,n}$. The dimension results discussed in the previous sections can be applied to show that, in many cases, ν_u has full dimension for all parameters u outside of a small set of exceptions (note that, in the definition of ν_u, we are skipping every k-th term only, so ν_u should be "almost as large" as μ_u). On the other hand, adapting a combinatorial method that goes back to Erdős [4] and has become known as the "Erdős-Kahane" argument, it is often possible to show that the Fourier transform

$$\hat{\eta}_u(\xi) = \int \exp(2\pi i \langle x, \xi \rangle)\, d\eta_u(x)$$

has a power decay, again outside of a small set of possible exceptions (because η_u is defined by keeping only every k-th term, these measures will have very small dimension and hence a very small power decay, but all that will matter is that it is positive).

Recall that the **Fourier dimension** of a measure η is defined as

$$\dim_{\mathsf{F}}(\eta) = \sup\{\sigma \geq 0 : \exists C, |\hat{\eta}(\xi)| \leq C|\xi|^{-\sigma/2}\}.$$

Absolute continuity then follows from the following general fact:

Theorem 6.1 *Let η, ν be Borel probability measures on $\mathbb{R}^d$.*

(i) *If $\dim_{\mathsf{H}} \nu + \dim_{\mathsf{F}} \eta > d$, then $\eta * \nu$ has an absolutely continuous density.*
(ii) *If $D_q\nu + \dim_{\mathsf{F}} \eta > d$ for some $q \in (1, 2]$, then $\eta * \nu$ has a density in L^q.*

The second part in the case $q = 2$ is a rather straightforward consequence of well-known identities relating D_2 to energies, and energies to the Fourier transform, while the first part follows from the second for $q = 2$ (or any other value of q); see [32, Lemma 2.1]. The second part for arbitrary values of $q \in (1, 2]$ is somewhat more involved, and relies on the Littlewood-Paley decomposition; it is proved in [33, Theorem 4.4] (where a version for $q \in (2, +\infty)$ is also established). The intuition behind the theorem is that convolving with a measure of positive Fourier dimension is a smoothing operation (positive Fourier dimension is a kind of "pseudo-randomness" indicator), which is enough to "upgrade" full or almost full dimension to absolute continuity.

We now indicate how to implement the above strategy for projections of planar self-similar measures μ. We need to assume that the IFS is homogeneous; this is to ensure that the measure we are projecting, and therefore also its projections, have the desired convolution structure.

Theorem 6.2 *Let $(f_i(x) = Sx + t_i)_{i\in\Lambda}$ be a homogeneous IFS on $\mathbb{R}^2$ satisfying the SSC and $\dim_{\mathsf{S}} \mathcal{F} > 1$. Then there exists a set $E \subset [0, \pi)$ of zero Hausdorff dimension, such that for all $\alpha \in [0, \pi) \setminus E$ the following holds:*

(i) *Let $\mu_p = \mu(\mathcal{F}, p)$. If $\dim_{\mathsf{H}} \mu_p > 1$, then $\Pi_\alpha\mu_p$ is absolutely continuous.*
(ii) *In the irrational rotation case, if $D_q\mu_p > 1$, then $\Pi_\alpha\mu_p$ has an L^q density.*
(iii) *Moreover, in the rational rotation case, if $\dim_{\mathsf{H}} \mu_p > 1$, then $\Pi_\alpha\mu_p$ has a density in L^q for some $q = q(\mathcal{F}, p, \alpha) > 1$.*

Note that, since $\mathcal{F}$ satisfies the SSC, there are explicit formulae for $\dim_{\mathsf{H}} \mu, D_q\mu$, and $\lim_{q\to 1^+} D_q\mu = \dim_{\mathsf{H}} \mu$, and the set of p to which the theorem applies is nonempty (it includes, for example, the natural weights). In particular, it follows from the second part that, in the irrational rotation case, if $\dim_{\mathsf{H}} \mu_p > 1$ then $\Pi_\alpha\mu_p$ has an L^q density for an explicit $q > 1$ that is independent of α and p. Also, if $D_q\mu_p < 1$, then $D_q\Pi_\alpha(\mu_p) < 1$, and $\Pi_\alpha(\mu_p)$ cannot have an L^q density, so the second part is sharp up to the endpoint. Thus we know a lot less about the density of the projections in the rational rotation case.

Sketch of proof First of all, by replacing $\mathcal{F}$ by $\mathcal{F}^k$ for suitable k, in the rational rotation case we may assume that S is a homothety, i.e. there are no rotations at all.

The self-similar measure μ_p is the distribution of the random sum $\sum_{n=1}^{\infty} X_n$, where $\mathbb{P}(X_n = S^n t_i) = p_i$, and the X_n are independent. Indeed, this measure is easily checked to satisfy the defining relation $\mu_p = \sum_{i \in \Lambda} p_i f_i \mu_p$. Since p is fixed in the proof we drop any explicit reference to it from now on.

As indicated above, let k be a large integer to be determined later, and let η, ν be the distribution of the random sums $\sum_{k|n} X_n$, $\sum_{k\nmid n} X_n$ respectively, so that $\mu = \eta * \nu$ and therefore $\Pi_\alpha \mu = \Pi_\alpha \eta * \Pi_\alpha \nu$. This fits with the general description above, since $\Pi_\alpha \mu$ is the distribution of $\sum_{n=1}^{\infty} X_{\alpha,n}$, where $\mathbb{P}(X_{\alpha,n} = \Pi_\alpha(S^n t_i)) = p_i$, and the $X_{\alpha,n}$ are independent, and likewise with $\Pi_\alpha \eta$, $\Pi_\alpha \nu$.

Both η, ν are again homogeneous self-similar measures of the same rotation type (irrational or no rotation), which also satisfy the SSC. Moreover, a direct calculation shows that

$$\dim_{\mathsf{H}} \nu = (1 - 1/k) \dim_{\mathsf{H}} \mu,$$

$$D_q \nu = (1 - 1/k) D_q \mu.$$

Consider first the irrational rotation case, and suppose p is such that $D_q \mu > 1$. Provided we chose k large enough, then also $D_q \nu > 1$. By Theorem 4.1, $D_q \Pi_\alpha \nu = 1$ for all α. On the other hand, a combinatorial argument similar to (although slightly more involved than) the classical Erdős' argument from [4], shows that $\dim_{\mathsf{F}} \Pi_\alpha \eta > 0$ outside of a possible exceptional set of zero Hausdorff dimension. See [33, Proposition 3.3] (this also holds in the no rotation case). Claim (ii) then follows from Theorem 6.1, and we have already seen that this implies (i) in the irrational rotation case.

The first claim in the no-rotations case follows in the same way, using Proposition 5.5 instead of Theorem 4.1. A priori this gives no information whatsoever about the densities (the reason being that Theorem 5.2 is about Hausdorff dimension and it is unknown if it holds for L^q dimension for any q). However, in [33, Theorem 5.1] we have shown that for any homogeneous self-similar measure τ, and in particular for $\tau = \Pi_\alpha \nu$ in the no-rotations case,

$$\lim_{q \to 1+} D_q(\tau) = \dim_{\mathsf{H}} \tau.$$

(This is immediate from the explicit formulae under the OSC, the point is that it holds regardless of overlaps.) Hence, if α is such that $\dim_{\mathsf{H}}(\Pi_\alpha \nu) = 1$ and $\dim_{\mathsf{F}}(\Pi_\alpha \eta) > 0$, there is a (non-explicit) $q > 1$ such that $D_q(\Pi_\alpha \nu) + \dim_{\mathsf{F}}(\Pi_\alpha \eta) > 1$. The third claim then follows again from Theorem 6.1. □

Using Lemma 4.2 once again, we conclude the proof of Theorem 2.1(iv) in the by now familiar way.

Unfortunately, the proof of Theorem 6.2 (and hence of Theorem 2.1(iv)) is completely non-effective. The reason is that it seems very hard to prove that a given projection of a self-similar measure has power Fourier decay, even though we know that all outside of a zero-dimensional set do!

7 Further Results

We briefly discuss projections of other natural classes of sets and measures. This section has some overlap with [7, Sections 8 and 9].

7.1 *Bernoulli Convolutions*

Given $\lambda \in (0, \frac{1}{2})$, the **Bernoulli convolution** ν_λ is the self-similar measure for the IFS $(\lambda x - 1, \lambda x + 1)$ with weights $(\frac{1}{2}, \frac{1}{2})$. Alternatively, ν_λ is the distribution of the random sum $\sum_{n=0}^{\infty} \pm\lambda^n$, where the signs are chosen independently with equal probabilities; this explains the name. When $\lambda \in (0, 1/2]$, the generating IFS satisfies the OSC and the measure ν_λ is well understood; however, for $\lambda \in (\frac{1}{2}, 1)$, surprisingly little is known. It is known since Erdős [3] that if λ^{-1} is a Pisot number (an algebraic integer larger than 1, all of whose algebraic conjugates are smaller than 1 in modulus), then ν_λ is singular. It is not known if there are any other $\lambda \in (\frac{1}{2}, 1)$ for which ν_λ is singular. Solomyak [35] proved that ν_λ is absolutely continuous with an L^2 density for almost all $\lambda \in (\frac{1}{2}, 1)$. This is a kind of Marstrand Theorem for a family of *nonlinear* projections. Using the method described in Sect. 6, the author proved in [32] that ν_λ is absolutely continuous for λ outside of a zero Hausdorff dimension set of exceptions, and in [33] we showed that, furthermore, outside this exceptional set, ν_λ has a density in L^q for some non-explicit $q = q(\lambda) > 1$. These results rely heavily on Theorem 5.2.

7.2 *Self-Similar Sets in Higher Dimension*

Much less is known about projections of self-similar sets in higher dimensions. In dimensions $d \geq 3$ there is no neat decomposition into "rational rotation" and "irrational rotation" cases. In particular, if the orthogonal parts of all the maps in the IFS coincide, then they cannot generate a dense subgroup of the orthogonal group – this is problematic for generalizing Theorems 4.1 and 6.2. Also, the lack of commutativity precludes approximation arguments such as Lemma 4.2. Nevertheless, the more flexible approach of [20] yields the following, see [20, Theorem 1.6 and Corollary 1.7].

Theorem 7.1 *Let $A \subset \mathbb{R}^d$, $d \geq 2$, be the attractor of the IFS $(\lambda_i O_i x + t_i)_{i\in\Lambda}$, where $\lambda_i \in (0,1)$, $O_i \in \mathbb{O}_d$ and $t_i \in \mathbb{R}^d$. Assume the SSC holds. Fix $1 \leq k < d$ and let $G_{d,k}$ denote the Grassmanian of k-dimensional subspaces of $\mathbb{R}^d$.*

Suppose that the action of the semigroup generated by the O_i on $G_{d,k}$ is transitive, that is, $\{O_{i_1}\cdots O_{i_n}\pi : i_j \in \Lambda, n \in \mathbb{N}\}$ is dense for some (and therefore all) $\pi \in G_{d,k}$. Then for all C^1 maps $g : \mathbb{R}^d \to \mathbb{R}^k$ without singular points,

$$\dim_{\mathsf{H}}(gA) = \min(\dim_{\mathsf{H}} A, k)$$

Note that in dimension $d = 2$, the transitivity condition is met precisely for self-similar sets of irrational type. Once again, this follows from a corresponding result for measures. Using an approximation argument, Farkas [11, Theorem 1.6] was able to remove the SSC assumption.

7.3 *Projections of Self-Affine Carpets*

If the maps f_i in an IFS are affine rather than similarities, the attractor is called a **self-affine set**. Dimension problems for self-affine sets are notoriously difficult, and almost nothing beyond the general results of Marstrand and others is known about their orthogonal projections, outside of some special classes known as *self-affine carpets*. Roughly speaking, a self-affine carpet is the attractor of an IFS of affine maps that map the unit square onto non-overlapping rectangles with some special pattern (generally speaking, it is required that when projecting these rectangles onto either the x or y-axes, there are either no overlaps or exact overlaps).

In [14], it was proved that under a suitable irrationality condition, for many self-affine carpets $A \subset \mathbb{R}^2$ it holds that $\dim_{\mathsf{H}} \Pi A = \min(\dim_{\mathsf{H}}(A), 1)$ for all projections Π other than the principal ones (which are always exceptional for carpets). The proof was based on ideas of [29] and did not extend to measures. Recently, based on the approach of [18], Ferguson, Fraser and Sahlsten [13] obtained the corresponding results for Bernoulli measures for the natural Markov partition for the $(x, y) \mapsto (px, qy)$ mod 1 toral endomorphism. This was extended to Gibbs measures by Almarza [1].

7.4 *Sums of Cantor Sets*

The arithmetic sum $A + B$ of two sets $A, B \subset \mathbb{R}^d$ is $\{x + y : x \in A, y \in B\}$. Up to an homothety, this is the projection of $A \times B$ under a 45 degree projection. More generally, the family $A + uB$ is a reparametrization of the projections of $A \times B$ (other than the horizontal projection). The methods discussed in the previous sections can be applied (with suitable modifications) to yield the following analog of Theorem 2.1. Following the terminology of [29], we say that the attractors of

$(\lambda_i x + t_i)$ and $(\lambda'_j x + t'_j)$ are **algebraically resonant** if $\log \lambda_i / \log \lambda'_j$ is rational for all i, j.

Theorem 7.2 *Let $A, B \subset \mathbb{R}$ be self-similar sets, and write $s = \dim_H(A) + \dim_H(B)$.*

(i) *If A and B are not algebraically resonant, then* $\dim_H(A + uB) = \min(s, 1)$ *for all* $u \in \mathbb{R} \setminus \{0\}$.
(ii) *If both A and B are given by algebraic parameters, then* $\{u : \dim_H(A + uB) < \min(s, 1)\}$ *is countable.*
(iii) *Without any assumptions, if* $s \leq 1$, *then* $\dim_H\{u : \dim_H(A + uB) < s\} = 0$.
(iv) *If* $s > 1$, *then* $\dim_H\{u : \mathcal{L}(A + uB) = 0\} = 0$.

Part (i) was proved in [29], (ii) and (iii) follow in a very similar manner to the corresponding claims in Theorem 2.1, and the last claim is [33, Theorem E]. Measure versions of these results exist, see [20, Theorem 1.4] and [33, Theorem D].

Much is known for sumsets beyond self-similar sets and measures. One of the first results in the area, due to Moreira, was a version of Theorem 7.2(i) for attractors of *nonlinear* IFSs (under standard regularity assumptions). See [2]. A general version of this (valid also for Gibbs measures) was obtained in [20, Theorem 1.4]. Furthermore, it follows from [20, Theorem 1.3] that if $A, B \subset [0, 1]$ are closed and invariant under $x \mapsto px \bmod 1$, $x \mapsto qx \bmod 1$ respectively, with $\log p / \log q \notin \mathbb{Q}$, then

$$\dim_H(A + uB) = \min(\dim_H(A) + \dim_H(B), 1) \quad \text{for all } u \in \mathbb{R} \setminus \{0\}.$$

For $u = 1$, this was another conjecture of Furstenberg.

7.5 *Projections of Random Sets and Measures*

There is a vast, and growing, literature on geometric properties of random sets and measures of Cantor type, including the behavior of their projections. We do no more than indicate some references for further reading. Generally speaking, there are two main strands of research in this area. One concerns random sets and measures that include deterministic ones as a special case. In this direction, Falconer and Jin [8, 9] investigated projections of random cascade measures (and related models) on self-similar sets, obtaining generalizations of several of the results we discussed. In [8], these results were applied in a clever way to study the dimension of linear sections of *deterministic* self-similar sets. The second line of research concerns sets and measures with a large degree of spatial independence; one of the most popular such models is fractal percolation, consisting in iteratively selecting random squares in the dyadic (or M-adic) grid. With stronger independence, one can typically say a lot more, for example proving positive Lebesgue measure, and even nonempty interior, for all projections simultaneously (compare with Theorem 2.1(iii)). See e.g.

[31] for results of this type for fractal percolation. A general approach to the study of measures with strong spatial independence was recently developed in [34]. We refer the reader to this paper for many further references and detailed statements.

References

1. J. Almarza, Projections of $(\times m, \times n)$-invariant Gibbs measures preserve dimension (2014, preprint). Available at http://arxiv.org/abs/1410.5086v1
2. C.G.T. de A. Moreira, Sums of regular Cantor sets, dynamics and applications to number theory. Period. Math. Hungar. **37**(1–3), 55–63 (1998). International Conference on Dimension and Dynamics (Miskolc, 1998)
3. P. Erdős, On a family of symmetric Bernoulli convolutions. Am. J. Math. **61**, 974–976 (1939)
4. P. Erdős, On the smoothness properties of a family of Bernoulli convolutions. Am. J. Math. **62**, 180–186 (1940)
5. K.I. Eroğlu, On planar self-similar sets with a dense set of rotations. Ann. Acad. Sci. Fenn. Math. **32**(2), 409–424 (2007)
6. K.J. Falconer, *Fractal Geometry*. Mathematical Foundations and Applications, 3rd edn. (Wiley, Chichester, 2014)
7. K.J. Falconer, J. Fraser, X. Jin, Sixty years of fractal projections (2014, preprint). Available at http://arxiv.org/abs/1411.3156v1
8. K.J. Falconer, X. Jin, Dimension conservation for self-similar sets and fractal percolation. Int. Math. Res. Notices (2014, accepted for publication). Available at http://arxiv.org/abs/1409.1882v2
9. K.J. Falconer, X. Jin, Exact dimensionality and projections of random self-similar measures and sets. J. Lond. Math. Soc. (2) **90**(2), 388–412 (2014)
10. A.-H. Fan, K.-S. Lau, H. Rao, Relationships between different dimensions of a measure. Monatsh. Math. **135**(3), 191–201 (2002)
11. Á. Farkas, Projections of self-similar sets with no separation condition. Isr. J. Math. (2014, to appear). Available at http://arxiv.org/abs/1307.2841v3
12. D.-J. Feng, H. Hu, Dimension theory of iterated function systems. Commun. Pure Appl. Math. **62**(11), 1435–1500 (2009)
13. A. Ferguson, J. Fraser, T. Sahlsten, Scaling scenery of $(\times m, \times n)$-invariant measures. Adv. Math. **268**, 564–602 (2015)
14. A. Ferguson, T. Jordan, P. Shmerkin, The Hausdorff dimension of the projections of self-affine carpets. Fund. Math. **209**(3), 193–213 (2010)
15. J. Fraser, P. Shmerkin, On the dimensions of a family of overlapping self-affine carpets. Ergod. Theory Dyn. Syst. (2014, accepted for publication). Available at http://arxiv.org/abs/1405.4919v2
16. A. Furman, On the multiplicative ergodic theorem for uniquely ergodic systems. Ann. Inst. H. Poincaré Probab. Stat. **33**(6), 797–815 (1997)
17. D. Galicer, S. Saglietti, P. Shmerkin, A. Yavicoli, L^q dimensions and projections of random measures. (preprint). Available at http://arxiv.org/abs/1504.04893
18. M. Hochman, On self-similar sets with overlaps and inverse theorems for entropy. Ann. Math. (2) **180**(2), 773–822 (2014)
19. M. Hochman, Self-similar sets, entropy and additive combinatorics, in *Geometry and Analysis of Fractals*, ed. by D.-J. Feng, K.-S. Lau. Springer Proceedings in Mathematics & Statistics, vol. 88 (Springer, Heidelberg, 2014), pp. 252–252
20. M. Hochman, P. Shmerkin, Local entropy averages and projections of fractal measures. Ann. Math. (2) **175**(3), 1001–1059 (2012)

21. X. Hu, S. James Taylor, Fractal properties of products and projections of measures in $\mathbf{R}^d$. Math. Proc. Camb. Philos. Soc. **115**(3), 527–544 (1994)
22. B.R. Hunt, V.Y. Kaloshin, How projections affect the dimension spectrum of fractal measures. Nonlinearity **10**(5), 1031–1046 (1997)
23. R. Kenyon, Projecting the one-dimensional Sierpinski gasket. Isr. J. Math. **97**, 221–238 (1997)
24. J.M. Marstrand, Some fundamental geometrical properties of plane sets of fractional dimensions. Proc. Lond. Math. Soc. (3) **4**, 257–302 (1954)
25. P. Mattila, *Geometry of Sets and Measures in Euclidean Spaces*. Volume 44 of Cambridge Studies in Advanced Mathematics (Cambridge University Press, Cambridge, 1995). Fractals and rectifiability
26. F. Nazarov, Y. Peres, P. Shmerkin, Convolutions of Cantor measures without resonance. Isr. J. Math. **187**, 93–116 (2012)
27. T. Orponen, On the distance sets of self-similar sets. Nonlinearity **25**(6), 1919–1929 (2012)
28. T. Orponen, On the packing measure of self-similar sets. Nonlinearity **26**(11), 2929–2934 (2013)
29. Y. Peres, P. Shmerkin, Resonance between Cantor sets. Ergod. Theory Dyn. Syst. **29**(1), 201–221 (2009)
30. Y. Peres, B. Solomyak, Existence of L^q dimensions and entropy dimension for self-conformal measures. Indiana Univ. Math. J. **49**(4), 1603–1621 (2000)
31. M. Rams, K. Simon, The geometry of fractal percolation, in *Geometry and Analysis of Fractals*, ed. by D.-J. Feng, K.-S. Lau. Springer Proceedings in Mathematics & Statistics, vol. 88 (Springer, Heidelberg, 2014), pp. 303–323
32. P. Shmerkin, On the exceptional set for absolute continuity of Bernoulli convolutions. Geom. Funct. Anal. **24**(3), 946–958 (2014)
33. P. Shmerkin, B. Solomyak, Absolute continuity of self-similar measures, their projections and convolutions. Trans. Am. Math. Soc. (2014, accepted for publication). Available at http://arxiv.org/abs/1406.0204v1
34. P. Shmerkin, V. Suomala, Spatially independent martingales, intersections, and applications (2014, preprint). Available at http://arxiv.org/abs/1409.6707v3
35. B. Solomyak, On the random series $\sum \pm\lambda^n$ (an Erdős problem). Ann. Math. (2) **142**(3), 611–625 (1995)

Part II
Self-Similar Fractals and Recurrent Structures

Dimension of the Graphs of the Weierstrass-Type Functions

Krzysztof Barański

Abstract We present a survey of results on dimension of the graphs of the Weierstrass-type functions on the real line.

Keywords Fractals • Hausdorff dimension • Weierstrass function

Mathematics Subject Classification (2000). Primary 28A78, 28A80; Secondary 37C45, 37C40, 37D25

1 Introduction

In this paper we consider continuous real functions of the form

$$f^{\phi}_{\lambda,b}(x) = \sum_{n=0}^{\infty} \lambda^n \phi(b^n x) \tag{1.1}$$

for $x \in \mathbb{R}$, where $b > 1$, $1/b < \lambda < 1$ and $\phi : \mathbb{R} \to \mathbb{R}$ is a non-constant, $\mathbb{Z}$-periodic, Lipschitz continuous, piecewise C^1 function. Probably the most famous function of that form is the *Weierstrass cosine function*

$$W_{\lambda,b}(x) = \sum_{n=0}^{\infty} \lambda^n \cos(2\pi b^n x),$$

introduced by Weierstrass in 1872 as one of the first examples of a continuous nowhere differentiable function on the real line. In fact, Weierstrass proved the non-differentiability for some values of the parameters λ, b, while the complete

Supported by the Polish NCN grant 2014/13/B/ST1/04551.

K. Barański (✉)
Institute of Mathematics, University of Warsaw, ul. Banacha 2, 02-097 Warszawa, Poland
e-mail: baranski@mimuw.edu.pl

C. Bandt et al. (eds.), *Fractal Geometry and Stochastics V*, Progress in Probability 70, DOI 10.1007/978-3-319-18660-3_5

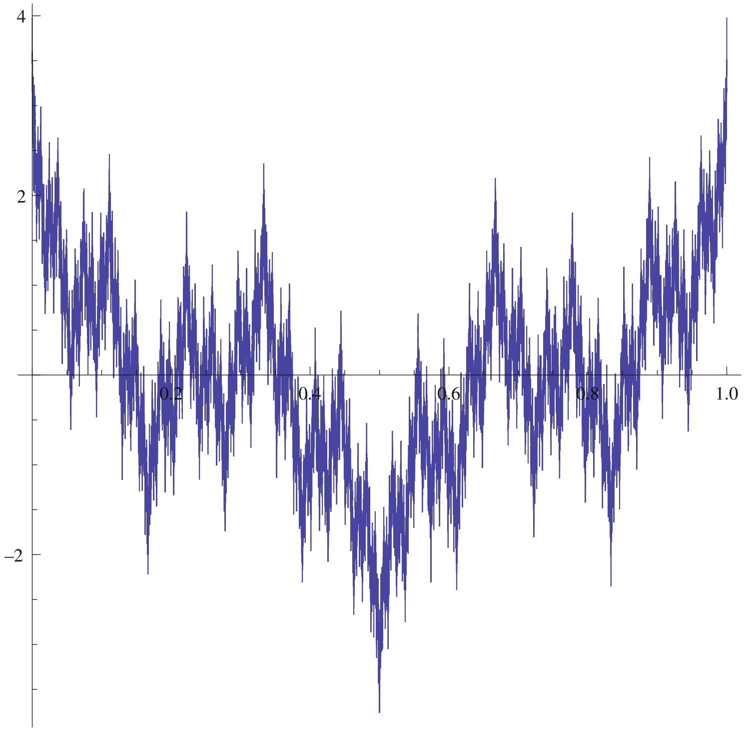


Fig. 1 The graph of the Weierstrass cosine nowhere differentiable function

proof for $b > 1$, $1/b < \lambda < 1$ was given by Hardy [16] in 1916. Later, starting from the work by Besicovitch and Ursell [8], the graphs of functions of the form (1.1) and related ones have been studied from a geometric point of view as fractal curves in the plane (Fig. 1).

In this paper we present a survey of recent results concerning various kinds of dimensions of the graphs of functions of the form (1.1).

Since

$$\lambda f^{\phi}_{\lambda,b}(bx) = f^{\phi}_{\lambda,b}(x) - \phi(x),$$

the graph of $f^{\phi}_{\lambda,b}$ exhibits a kind of approximate self-affinity with scales λ and $1/b$, which suggests a candidate for its dimension to be

$$D = 2 + \frac{\log \lambda}{\log b}.$$

We will see that this is indeed the case for the box dimension of the graph of $f^{\phi}_{\lambda,b}$ (unless it is a piecewise C^1 function with the graph of dimension 1), see Theorem 2.4. For the Hausdorff dimension, the situation is not so simple – we know some general lower estimates by constants smaller than D (see (3.1) and

Theorem 3.4), while the Hausdorff dimension of the graph is known to be equal to D only in some concrete cases (see Theorems 3.5 and 3.6), and for integer b and generic smooth function ϕ (see Theorem 4.2). On the other hand, we do not know any example of a function of the form (1.1), where the Hausdorff dimension of the graph is smaller than D.

Let us note that if b is an integer, then the graph of a function $f^{\phi}_{\lambda,b}$ of the form (1.1) is an invariant repeller for the expanding dynamical system

$$\Phi : \mathbb{R}/\mathbb{Z} \times \mathbb{R} \to \mathbb{R}/\mathbb{Z} \times \mathbb{R}, \qquad \Phi(x, y) = \left(bx \,(\mathrm{mod}\, 1),\ \frac{y - \phi(x)}{\lambda} \right) \tag{1.2}$$

with two different positive Lyapunov exponents $-\log \lambda < \log b$, which allows to use the methods of ergodic theory of smooth dynamical systems. In this case the graph of $f^{\phi}_{\lambda,b}$ is the common fractal boundary between the basins of attraction to (vertical) $+\infty$ and $-\infty$ on the cylinder $\mathbb{R}/\mathbb{Z}$ (see Fig. 2). Alternatively, the system

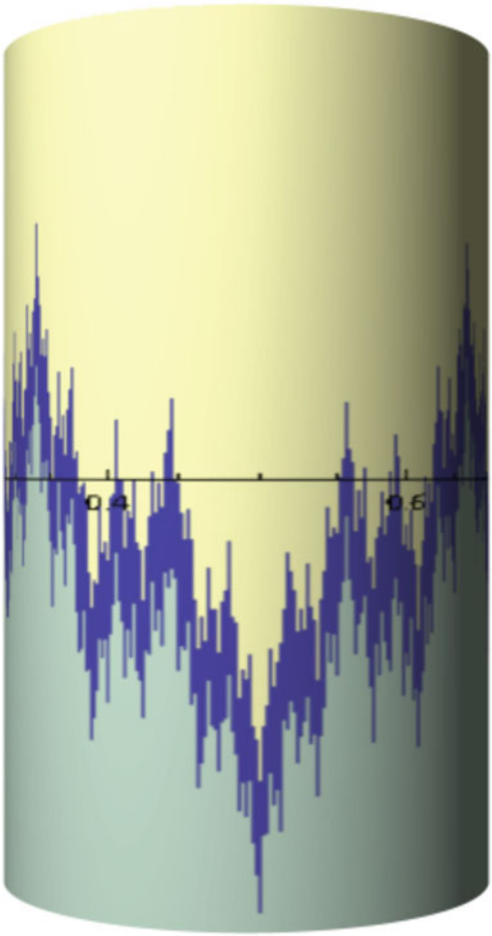


Fig. 2 The graph of the Weierstrass function as the boundary between two attracting basins on the cylinder

can be treated as a nonlinear iterated function system (IFS) on $[0, 1) \times \mathbb{R}$ composed of the maps

$$S_i(x, y) = \left(\frac{x}{b} + \frac{i}{b},\ \lambda y + \phi\left(\frac{x}{b} + \frac{i}{b}\right)\right), \qquad i = 0, \ldots, b-1.$$

Some results presented in this paper are valid also for a more general class of functions

$$f_{\lambda,b}^{\phi,\Theta}(x) = \sum_{n=0}^{\infty} \lambda^n \phi(b^n x + \theta_n), \tag{1.3}$$

where $\Theta = (\theta_1, \theta_2, \ldots)$ for $\theta_n \in \mathbb{R}$ is a sequence of *phases* (with the previous assumptions on λ, b and ϕ).

We will consider the Hausdorff, packing and box dimension denoted, respectively, by $\dim_H$, $\dim_P$ and $\dim_B$. The upper and lower box dimension will be denoted, respectively, by $\overline{\dim}_B$ and $\underline{\dim}_B$. For an unbounded set, the (upper, lower) box dimension is defined as the supremum of (upper, lower) box dimensions of its bounded subsets.

For the definitions of the considered dimensions and their basic properties we refer to [14, 28]. We only note that for a set $X \subset \mathbb{R}^k$ we have

$$\dim_H(X) \le \underline{\dim}_B(X) \le \overline{\dim}_B(X)$$

and

$$\dim_H(X) \le \dim_P(X) \le \overline{\dim}_B(X).$$

The plan of the paper is as follows. In Sect. 2 we determine the box and packing dimension of the graphs of functions of the form (1.1). Results on the Hausdorff dimension are presented in Sects. 3–4. In Sect. 5 we deal with a randomization of functions of the form (1.3). Additional issues (complex extension of the Weierstrass cosine function, non-exponential sequences of scalings) are treated in Sects. 6–7.

Note that the quoted results are not necessarily presented in the chronological order and the formulation can be different from the original. Due to lack of space, the proofs are generally not included and the reader is referred to original articles.

There are a number of related issues which are not discussed in this paper (e.g. the case $\lambda = b$, wider classes of functions ϕ, dimension of the graphs of self-affine functions). The reader can find some information on these questions in the works included in the bibliography and the references therein.

2 Local Oscillations, Hölder Condition and Box Dimension

To determine the box dimension of the graphs of the considered functions, we examine their local oscillations in terms of the Hölder condition. By I we denote a non-trivial (not necessarily bounded) interval in $\mathbb{R}$.

Definition 2.1 We say that a function $f : I \to \mathbb{R}$ is *Hölder continuous* with exponent $\beta > 0$, if there exist $c, \delta > 0$ such that

$$|f(x) - f(y)| \leq c|x - y|^{\beta}$$

for every $x, y \in I$ such that $|x - y| < \delta$. Hölder continuous functions with exponent 1 are called *Lipschitz continuous* (with the *Lipschitz constant* c).

We say that f satisfies the *lower Hölder condition* with exponent $\beta > 0$, if there exist $\tilde{c}, \tilde{\delta} > 0$ such that the *oscillation*

$$\operatorname{osc}_J(f) = \sup_J f - \inf_J f$$

of f on every interval $J \subset I$ with $|J| < \tilde{\delta}$ satisfies

$$\operatorname{osc}_J(f) \geq \tilde{c}|J|^{\beta}$$

(where $|\cdot|$ denotes the length).

Note that the lower Hölder condition with exponent $\beta \in (0, 1)$ implies non-differentiability (i.e. non-existence of a finite derivative) of the function at every point.

The following proposition follows directly from the definitions of the upper and lower box dimension.

Proposition 2.2 *If $f : I \to \mathbb{R}$ is Hölder continuous with exponent $\beta \in (0, 1]$, then*

$$\overline{\dim}_B(\operatorname{graph} f) \leq 2 - \beta.$$

If a continuous function $f : I \to \mathbb{R}$ satisfies the lower Hölder condition with exponent $\beta \in (0, 1]$, then

$$\underline{\dim}_B(\operatorname{graph} f) \geq 2 - \beta.$$

Let

$$\alpha = -\frac{\log \lambda}{\log b} = 2 - D.$$

Note that by definition, $\alpha \in (0, 1)$ and $\lambda = b^{-\alpha}$. The following upper estimate of the box dimension of the graphs of the considered functions is a consequence of Proposition 2.2.

Proposition 2.3 *Every function $f_{\lambda,b}^{\phi,\Theta}$ of the form* (1.3) *is Hölder continuous with exponent α, and hence $\overline{\dim}_B(\text{graph } f_{\lambda,b}^{\phi,\Theta}) \leq D$.*

Proof Let c be a Lipschitz constant of ϕ. Take $x, y \in I$ such that $0 < |x - y| \leq 1$. Then, choosing $N \in \mathbb{N}$ with $1/b^N < |x - y| \leq 1/b^{N-1}$, we have

$$|f_{\lambda,b}^{\phi,\Theta}(x) - f_{\lambda,b}^{\phi,\Theta}(y)| \leq c|x-y| \sum_{n=0}^{N-1} (\lambda b)^n + 2 \max \phi \sum_{n=N}^{\infty} \lambda^n$$
$$< \left(\frac{cb}{\lambda b - 1} + \frac{2 \max \phi}{1 - \lambda} \right) \lambda^N < \left(\frac{cb}{\lambda b - 1} + \frac{2 \max \phi}{1 - \lambda} \right) |x - y|^{\alpha}.$$

□

One cannot expect a non-trivial lower estimate for the dimension of the graph, which holds for every function under consideration. Indeed, if

$$\phi(x) = g(x) - \lambda g(bx)$$

for an integer $b > 1$ and a $\mathbb{Z}$-periodic, Lipschitz continuous, piecewise C^1 function g, then $f_{\lambda,b}^{\phi}$ has the form (1.1) and $f_{\lambda,b}^{\phi} = g$, so its graph is a piecewise smooth curve of dimension 1. However, for functions of the form (1.1), the case of a piecewise C^1 curve is the only possible exception, when the box dimension of the graph is smaller than D. The following fact is a consequence of a result by Hu and Lau [18].

Theorem 2.4 *For every function $f_{\lambda,b}^{\phi}$ of the form* (1.1)*, exactly one of the two following possibilities holds.*

(a) *$f_{\lambda,b}^{\phi}$ is piecewise C^1 (and hence the dimension of its graph is 1).*
(b) *$f_{\lambda,b}^{\phi}$ satisfies the lower Hölder condition with exponent α (in particular it is nowhere differentiable) and $\dim_B(\text{graph } f_{\lambda,b}^{\phi}) = D$.*

Proof Adding a constant to $f_{\lambda,b}^{\phi}$, we can assume $\phi(0) = 0$. In [18, Theorem 4.1] it is proved that in this case, if the *Weierstrass–Mandelbrot function*

$$V(x) = \sum_{n=-\infty}^{\infty} \lambda^n \phi(b^n x) = f_{\lambda,b}^{\phi} + \sum_{n=1}^{\infty} \frac{1}{\lambda^n} \phi\left(\frac{x}{b^n}\right)$$

is not identically zero, then $f_{\lambda,b}^{\phi}$ satisfies the lower Hölder condition with exponent α. Hence, if $V \not\equiv 0$, then we can use Propositions 2.2 and 2.3 to obtain the assertion (b). On the other hand, $\sum_{n=1}^{\infty}(1/\lambda^n)\phi(x/b^n)$ is a piecewise C^1 function, so the condition $V \equiv 0$ implies that $f_{\lambda,b}^{\phi}$ is piecewise C^1, which is the case (a). □

A consequence of Theorem 2.4 is that the graphs of functions of the form (1.1) have packing dimension equal to box dimension.

Proposition 2.5 *For every function* $f^{\phi}_{\lambda,b}$ *of the form* (1.1)*, we have*

$$\dim_P(\text{graph } f^{\phi}_{\lambda,b}) = \dim_B(\text{graph } f^{\phi}_{\lambda,b}).$$

Proof We can assume that we are in the case (b) of Theorem 2.4, i.e. $f^{\phi}_{\lambda,b}$ satisfies the lower Hölder condition with exponent α and $\dim_B(\text{graph } f^{\phi}_{\lambda,b}) = D$. It is a general fact (see [14, Corollary 3.9]), that for every compact set $X \subset \mathbb{R}^k$, if

$$\overline{\dim}_B(X \cap U) = \overline{\dim}_B(X)$$

for every open set U intersecting X, then $\dim_P(X) = \overline{\dim}_B(X)$. To prove the proposition, we check this condition for $X = \text{graph } f^{\phi}_{\lambda,b}|_I$, where I is an arbitrary non-trivial compact interval in R.

By the continuity of $f^{\phi}_{\lambda,b}$, for an open set U intersecting graph $f^{\phi}_{\lambda,b}|_I$, we can take a non-trivial interval $J \subset I$ such that graph $f^{\phi}_{\lambda,b}|_J \subset$ graph $f^{\phi}_{\lambda,b}|_I \cap U$. Since $f^{\phi}_{\lambda,b}|_J$ satisfies the lower Hölder condition with exponent α, Proposition 2.2 implies

$$D \le \overline{\dim}_B(\text{graph } f^{\phi}_{\lambda,b}|_J) \le \overline{\dim}_B(\text{graph } f^{\phi}_{\lambda,b}|_I \cap U) \le \overline{\dim}_B(\text{graph } f^{\phi}_{\lambda,b}|_I) \le D,$$

which ends the proof. □

In particular, the Weierstrass cosine function $W_{\lambda,b}$ satisfies the lower Hölder condition with exponent α and

$$\dim_P(\text{graph } W_{\lambda,b}) = \dim_B(\text{graph } W_{\lambda,b}) = D$$

for $b > 1$, $1/b < \lambda < 1$. Similar results for various classes of functions ϕ were obtained, among others, by Kaplan, Mallet-Paret and Yorke [24], Rezakhanlou [34], Przytycki and Urbański [33] and Bousch and Heurteaux [9].

In [17], Heurteaux generalized the above results to the case of functions of the form (1.3) with transcendental b.

Theorem 2.6 *Every function* $f^{\phi,\Theta}_{\lambda,b}$ *of the form* (1.3)*, where* b *is a transcendental number, satisfies the lower Hölder condition with exponent* α (*in particular it is nowhere differentiable*)*. Moreover,*

$$\dim_P(\text{graph } f^{\phi,\Theta}_{\lambda,b}) = \dim_B(\text{graph } f^{\phi,\Theta}_{\lambda,b}) = D.$$

3 Hausdorff Dimension

The question of determining the Hausdorff dimension of the graphs of the considered functions is much more delicate and far from being completely solved. Since the upper bound $\dim_H \le \overline{\dim}_B \le D$ is known, one looks for suitable lower estimates. A standard tool is to analyse local properties of a finite Borel measure on the graph.

Definition 3.1 Let μ be a finite Borel measure in a metric space X. The *upper* and *lower local dimension* of μ at a point $x \in X$ are defined, respectively, as

$$\overline{\dim}\ \mu(x) = \limsup_{r \to 0+} \frac{\log \mu(B_r(x))}{\log r}, \qquad \underline{\dim}\ \mu(x) = \liminf_{r \to 0+} \frac{\log \mu(B_r(x))}{\log r},$$

where $B_r(x)$ denotes the ball of radius r centered at x. If for some d the upper and lower local dimensions of μ at x coincide and are equal to d for μ-almost every x, then we say that μ has *local dimension* d and write $\dim \mu = d$. Such measures are also called *exact-dimensional*.

Estimating the Hausdorff dimension of a set, it is standard to use the following fact (see [14, 28]).

Lemma 3.2 *If for some $d > 0$ we have $\underline{\dim}\ \mu(x) \geq d$ for μ-almost every x, then every Borel set of positive measure μ has Hausdorff dimension at least d. In particular, this holds if* $\dim \mu \geq d$.

In [33], using Lemma 3.2 for the lift of the Lebesgue measure on $[0, 1]$ to the graph of the function, Przytycki and Urbański proved the following.

Theorem 3.3 *If $f : I \to \mathbb{R}$ is Hölder continuous with exponent $\beta \in (0, 1)$ and satisfies the lower Hölder condition with exponent β, then*

$$\dim_H(\text{graph } f) > C > 1,$$

where C depends only on β and constants c, $\tilde{c}$ in Definition 2.1.

This together with Proposition 2.3 and Theorem 2.4 implies that for every function $f_{\lambda,b}^{\phi}$ of the form (1.1), if $f_{\lambda,b}^{\phi}$ is not piecewise C^1, then

$$\dim_H(\text{graph } f_{\lambda,b}^{\phi}) > 1. \tag{3.1}$$

Better estimates can be obtained for large b, even in the presence of phases, as shown by Mauldin and Williams [29].

Theorem 3.4 *For every function $f_{\lambda,b}^{\phi,\Theta}$ of the form* (1.3)*, there exists a constant $B >$ 0 depending only on λ and ϕ, such that*

$$\dim_H(\text{graph } f_{\lambda,b}^{\phi,\Theta}) > D - \frac{B}{\log b}$$

for every sufficiently large b.

The result was obtained by using Lemma 3.2 for the lift of a measure supported on a suitable Cantor set in $[0, 1]$ to the graph of the function.

The first example of a function of the form (1.1) with the graph of Hausdorff dimension equal to D was given by Ledrappier [25], who proved the following result, using the theory of invariant measures for non-uniformly hyperbolic smooth dynamical systems (Pesin theory) on manifolds [26].

Theorem 3.5 *For* $\phi(x) = \text{dist}(x, \mathbb{Z})$ (*the* sawtooth function) *and* $b = 2$,

$$\dim_H(\text{graph } f^{\phi}_{\lambda,2}) = D$$

for Lebesgue almost all $\lambda \in (1/2, 1)$.

In fact, the assertion holds provided the infinite Bernoulli convolution $\sum_{n=0}^{\infty} \pm 1/(2\lambda)^n$, with $\pm$ chosen independently with probabilities $(1/2, 1/2)$, has absolutely continuous distribution. As proved by Solomyak [39], the condition is fulfilled for almost all $\lambda \in (1/2, 1)$. By a recent result by Shmerkin [38], in fact it holds for all $\lambda \in (1/2, 1)$ except of a set of Hausdorff dimension 0.

In [40], Solomyak generalized the result from Theorem 3.5 to the case of some functions ϕ with discontinuous derivative (*nonlinear sawtooth functions*).

For the Weierstrass cosine function $W_{\lambda,b}$, the conjecture that the Hausdorff dimension of its graph is equal to D was formulated by Mandelbrot [27] in 1977 (see also [7]) and then repeated in a number of subsequent papers. Recently, Bárány, Romanowska and the author [5] proved the following result, showing that the conjecture is true for every nonzero integer b and a large set of parameters λ.

Theorem 3.6 *For every integer* $b > 1$ *there exist* $\lambda_b, \tilde{\lambda}_b \in (1/b, 1)$, *such that for every* $\lambda \in (\lambda_b, 1)$ *and Lebesgue almost every* $\lambda \in (\tilde{\lambda}_b, 1)$, *we have* $\dim \mu_{\lambda,b} = D$, *where* $\mu_{\lambda,b}$ *is the lift of the Lebesgue measure on* $[0, 1]$ *to* graph $W_{\lambda,b}$. *In particular,*

$$\dim_H(\text{graph } W_{\lambda,b}) = D$$

for every $\lambda \in (\lambda_b, 1)$ *and almost every* $\lambda \in (\tilde{\lambda}_b, 1)$. *We have*

$$\lambda_2 < 0.9352, \quad \lambda_3 < 0.7269, \quad \lambda_4 < 0.6083, \quad \lambda_b < 0.5448 \quad \text{for every } b \geq 5,$$
$$\tilde{\lambda}_2 < 0.81, \quad \tilde{\lambda}_3 < 0.55, \quad \tilde{\lambda}_4 < 0.44, \quad \tilde{\lambda}_b < 1.04/\sqrt{b} \quad \text{for every } b \geq 5$$

and

$$\lambda_b \to \frac{1}{\pi}, \quad \tilde{\lambda}_b \sqrt{b} \to \frac{1}{\sqrt{\pi}} \quad \text{as} \quad b \to \infty.$$

The proof uses the Ledrappier–Young theory from [26], Tsujii's results [41] on the Sinai–Bowen–Ruelle (SBR) measure for some smooth Anosov endomorphisms of the cylinder and the Peres–Solomyak transversality methods developed under the study of infinite Bernoulli convolutions (see e.g. [31, 32, 40]).

4 Dimension of Graphs of Generic Functions

In mathematics there are a number of definitions of a generic (typical) property. A *topologically generic* set in a space X is a set containing an open and dense set in X, or a *residual* set (containing a countable intersection of open and dense sets in X). A *measure-theoretic generic* set in $\mathbb{R}^k$ is a set of full Lebesgue measure. We use the following infinite-dimensional analogue of this property, which is called *prevalence* (see e.g. [30]).

Definition 4.1 A Borel set E in a real vector space V is *prevalent*, if there exists a finite set $\{v_1, \ldots, v_k\} \subset V$ (called the *probe set*), such that for every $v \in V$, one has $v + \sum_{j=1}^{k} t_j v_j \in E$ for Lebesgue almost every $(t_1, \ldots, t_k) \in \mathbb{R}^k$. A non-Borel subset of V is prevalent, if it contains a Borel prevalent set.

The topological and measure-theoretical genericity need not coincide. In fact, a topologically typical (residual) continuous function on $[0, 1]$ is nowhere differentiable (this follows from the Baire Theorem, see [1]) and has the graph of lower box dimension 1 (see [22]) and packing dimension 2 (see [19]), while a measure-theoretic typical (prevalent) continuous function on $[0, 1]$ is nowhere differentiable (see [20]) and has the graph of Hausdorff dimension 2 (see [15]). In [12] (see also [36]), using the wavelet technique, it was proved that functions with graphs of Hausdorff dimension $2 - \beta$ are prevalent within the space of Hölder continuous functions on $\mathbb{R}$ with given exponent $\beta \in (0, 1)$.

In [5], Bárány, Romanowska and the author proved that for functions $f_{\lambda,b}^{\phi}$ of the form (1.1) with integer b, the Hausdorff dimension of graph $f_{\lambda,b}^{\phi}$ is equal to D both for topologically and measure-theoretic typical C^3 function ϕ. To formulate the result precisely, consider the space $C^r(\mathbb{R}/\mathbb{Z})$, for $r = 3, 4, \ldots, \infty$, of $\mathbb{Z}$-periodic C^r real functions on $\mathbb{R}$, treated as functions on $\mathbb{R}/\mathbb{Z}$. For $b > 1$ let

$$\mathcal{F}_b = \{(\lambda, \phi) \in (1/b, 1) \times C^3(\mathbb{R}/\mathbb{Z}) : \dim \mu_{\lambda,b}^{\phi} = D\},$$

where $\mu_{\lambda,b}^{\phi}$ is the lift of the Lebesgue measure on $[0, 1]$ to graph $f_{\lambda,b}^{\phi}$. Recall that

$$\dim_H(\text{graph } f_{\lambda,b}^{\phi}) = D \qquad \text{for every } (\lambda, \phi) \in \mathcal{F}_b.$$

For $\lambda \in (1/b, 1)$, let

$$\mathcal{E}_{\lambda,b} = \{\phi \in C^3(\mathbb{R}/\mathbb{Z}) : (\lambda, \phi) \in \operatorname{int} \mathcal{F}_b\},$$

where int denotes the interior with respect to the product of the Euclidean and C^3 topology in $(1/b, 1) \times C^3(\mathbb{R}/\mathbb{Z})$. In [5], the following result was proved.

Theorem 4.2 *For every integer $b > 1$ and $\lambda \in (1/b, 1)$, the set $\mathcal{E}_{\lambda,b}$ is prevalent as a subset of $C^3(\mathbb{R}/\mathbb{Z})$, with a probe set contained in $C^\infty(\mathbb{R}/\mathbb{Z})$. Consequently, for every $r = 3, 4, \ldots, \infty$, the set $\mathcal{E}_{\lambda,b}$ is an open and dense subset of $C^r(\mathbb{R}/\mathbb{Z})$, and the set $\mathcal{F}_b$ contains an open and dense subset of $(1/b, 1) \times C^3(\mathbb{R}/\mathbb{Z})$.*

Similarly as for Theorem 3.6, the proof is based on the Ledrappier–Young theory from [26] and a result by Tsujii [41] on the generic absolute continuity of the SBR measure for some smooth Anosov endomorphisms of the cylinder.

5 Randomization

It is a well-known fact that introducing some additional parameters or stochastics to a system can sometimes help to answer questions which could not be solved in a standard setup. In studying dimension of the graphs of functions of the form (1.3), a number of results were obtained by randomizing suitable parameters.

Concerning the box dimension, Heurteaux [17] proved the following.

Theorem 5.1 *Let $f_{\lambda,b}^{\phi,\Theta}$ be a function of the form* (1.3)*. If one considers the phases θ_n as independent random variables with uniform distribution on* $[0, 1]$*, then almost surely, $f_{\lambda,b}^{\phi,\Theta}$ satisfies the lower Hölder condition with exponent $\alpha + \varepsilon$, for arbitrarily small $\varepsilon > 0$* (*in particular it is nowhere differentiable*)*. Moreover,*

$$\dim_P(\text{graph } f_{\lambda,b}^{\phi,\Theta}) = \dim_B(\text{graph } f_{\lambda,b}^{\phi,\Theta}) = D \quad \text{almost surely.}$$

An analogous result on the Hausdorff dimension can be obtained with stronger assumptions on the function ϕ, as proved by Hunt [21].

Theorem 5.2 *Let $f_{\lambda,b}^{\phi,\Theta}$ be a function of the form* (1.3)*, where ϕ is a C^∞ function with bounded set of orders of all its critical points. If one considers the phases θ_n as independent random variables with uniform distribution on* $[0, 1]$*, then*

$$\dim_H(\text{graph } f_{\lambda,b}^{\phi,\Theta}) = D \quad \text{almost surely.}$$

This includes the case, when ϕ is real-analytic, in particular (*for $\phi(x) = \cos(2\pi x)$*)*, when $f_{\lambda,b}^{\phi,\Theta}$ is the Weierstrass cosine function with phases θ_n.*

Similar results of that kind were obtained by Romanowska [35], using randomization of the parameter λ.

6 Complex Extension of the Weierstrass Function

It is interesting to notice that if b is an integer, then the Weierstrass cosine function $W_{\lambda,b}$ is the real part of the lacunary (Hadamard gaps) complex power series

$$w(z) = \sum_{n=0}^{\infty} \lambda^n z^{b^n}, \qquad z \in \mathbb{C}, \ |z| \leq 1$$

on the unit circle $\{|z| = 1\}$. In particular, $W_{\lambda,b}$ has a harmonic extension to the unit disc. This approach was already used by Hardy [16] to prove the non-differentiability of $W_{\lambda,b}$ in this case. The study of the boundary behaviour of the holomorphic map w is itself an interesting question. Salem and Zygmund [37] and Kahane, Weiss and Weiss [23] proved that for given b, if λ is sufficiently close to 1, then the image of the unit circle under w is a Peano curve, i.e. it covers an open subset of the plane. In [2], the author showed that in this case the box dimension of the graph of the function

$$x \mapsto \left(\sum_{n=0}^{\infty} \lambda^n \cos(2\pi b^n x), \ \sum_{n=0}^{\infty} \lambda^n \sin(2\pi b^n x) \right)$$

(which is a subset of $\mathbb{R}^3$), is equal to $3 - 2\alpha$. Moreover, the author [3] (see also Belov [6]) showed that for given b, if λ is sufficiently close to 1, then the map w does not preserve (forwardly) Borel sets on the unit circle. The boundary behaviour of w from a topological point of view was studied by Dong, Lau and Liu [13] (Fig. 3).

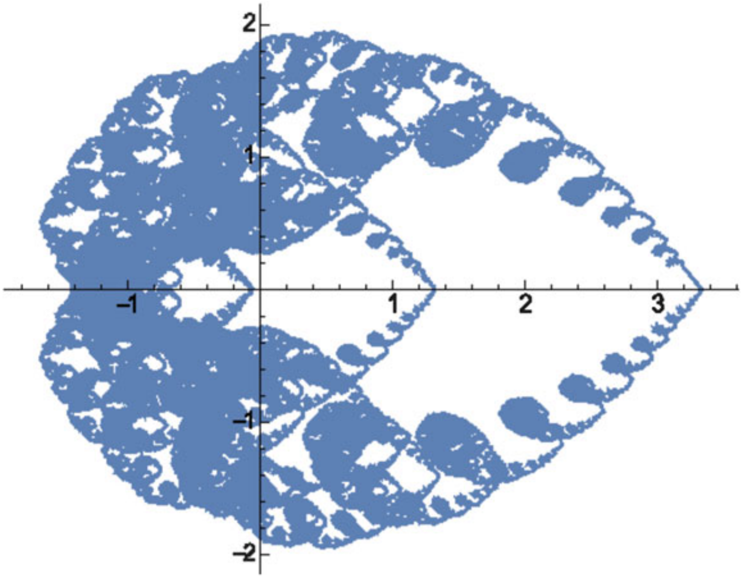


Fig. 3 The image of the unit circle under the map w

7 Other Sequences of Scalings

It is natural to study a generalization of functions of the forms (1.1) and (1.3), replacing λ^n, b^n by another sequences of scales λ_n, b_n, which are not exponential. More precisely, one can consider functions of the form

$$f(x) = \sum_{n=0}^{\infty} \lambda_n \phi(b_n x + \theta_n) \tag{7.1}$$

for $\lambda_n, b_n > 0$, $\sum_{n=0}^{\infty} \lambda_n < \infty$, $b_{n+1} > b_n$, $b_n \to \infty$ and a non-constant, $\mathbb{Z}$-periodic, Lipschitz continuous, piecewise C^1 function $\phi : \mathbb{R} \to \mathbb{R}$.

It turns out that the case of rapidly (faster than exponential) growing scales $1/\lambda_n$, b_n is easier to handle than the exponential one. In 1937, Besicovitch and Ursell [8] considered this case rather than the exponential one, and showed that for the sawtooth function $\phi(x) = \operatorname{dist}(x, \mathbb{Z})$, if $\lambda_n = b_n^{-\alpha}$ for some $\alpha \in (0, 1)$ and b_{n+1}/b_n tends to ∞ sufficiently slowly, then $\dim_H(\operatorname{graph} f) = 2 - \alpha$. The Hausdorff, upper and lower box dimension of the graphs of functions of the general form (7.1) for rapidly growing scales $1/\lambda_n$, b_n was computed by Carvalho [10] and the author [4]. More precisely, the following result was proved in [4].

Theorem 7.1 *For every function f of the form (7.1), if $\lambda_{n+1}/\lambda_n \to 0$, $b_{n+1}/b_n \to \infty$ as $n \to \infty$, then*

$$\dim_H(\operatorname{graph} f) = \underline{\dim}_B(\operatorname{graph} f) = 1 + \liminf_{n\to\infty} \frac{\log^+ d_n}{\log(b_{n+1} d_n / d_{n+1})},$$

$$\overline{\dim}_B(\operatorname{graph} f) = 1 + \limsup_{n\to\infty} \frac{\log^+ d_n}{\log b_n},$$

where $\log^+ = \max(\log, 0)$ *and* $d_n = \lambda_1 b_1 + \cdots + \lambda_n b_n$*.*

If additionally, $\lambda_n = b_n^{-\alpha}$ *for some* $\alpha \in (0, 1)$ *and* $\log b_{n+1} / \log b_n \to 1$*, then*

$$\dim_H(\operatorname{graph} f) = \dim_P(\operatorname{graph} f) = \dim_B(\operatorname{graph} f) = 2 - \alpha.$$

In particular, this shows that in the case of rapidly growing scales, the dimensions need not coincide. In fact, in [4] it is shown that for every $H, B \in [1, 2]$ with $H \le B$ one can find a function f satisfying the assumptions of Theorem 7.1, such that

$$\dim_H(\operatorname{graph} f) = \underline{\dim}_B(\operatorname{graph} f) = H, \qquad \overline{\dim}_B(\operatorname{graph} f) = B.$$

The case when the scales $1/\lambda_n$, b_n grow slower than exponentially is much more difficult and almost nothing is known about the dimension of the graph of f. Some work was done for

$$R(x) = \sum_{n=1}^{\infty} \frac{1}{n^2} \sin(2\pi n^2 x)$$

(called the *Riemann example*) and similar functions. In particular, Chamizo [11] determined the box dimension of the graph of R to be equal to $5/4$.

References

1. S. Banach, Über die Baire'sche Kategorie gewisser Funktionenmengen. Studia Math. **3**, 174–179 (1931)
2. K. Barański, On the complexification of the Weierstrass non-differentiable function. Ann. Acad. Sci. Fenn. Math. **27**(2), 325–339 (2002)
3. K. Barański, On some lacunary power series. Mich. Math. J. **54**(1), 65–79 (2006)
4. K. Barański, On the dimension of graphs of Weierstrass-type functions with rapidly growing frequencies. Nonlinearity **25**(1), 193–209 (2012)
5. K. Barański, B. Bárány, J. Romanowska, On the dimension of the graph of the classical Weierstrass function. Adv. Math. **265**, 32–59 (2014)
6. A.S. Belov, On the Salem and Zygmund problem with respect to the smoothness of an analytic function that generates a Peano curve. Mat. Sb. **181**(8), 1048–1060 (1990); translation in Math. USSR-Sb. **70**(2), 485–497 (1991)
7. M.V. Berry, Z.V. Lewis, On the Weierstrass-Mandelbrot fractal function. Proc. R. Soc. Lond. Ser. A **370**(1743), 459–484 (1980)
8. A.S. Besicovitch, H.D. Ursell, Sets of fractional dimensions (V): on dimensional numbers of some continuous curves. J. Lond. Math. Soc. **1**(1), 18–25 (1937)
9. T. Bousch, Y. Heurteaux, Caloric measure on domains bounded by Weierstrass-type graphs. Ann. Acad. Sci. Fenn. Math. **25**(2), 501–522 (2000)
10. A. Carvalho, Hausdorff dimension of scale-sparse Weierstrass-type functions. Fund. Math. **213**(1), 1–13 (2011)
11. F. Chamizo, Automorphic forms and differentiability properties. Trans. Am. Math. Soc. **356**(5), 1909–1935 (2004)
12. M. Clausel, S. Nicolay, Some prevalent results about strongly monoHölder functions. Nonlinearity **23**(9), 2101–2116 (2010)
13. X.-H. Dong, K.-S. Lau, J.-C. Liu, Cantor boundary behavior of analytic functions. Adv. Math. **232**, 543–570 (2013)
14. K. Falconer, *Fractal Geometry. Mathematical Foundations and Applications*, 2nd edn. (Wiley, Hoboken, 2003)
15. J.M. Fraser, J.T. Hyde, The Hausdorff dimension of graphs of prevalent continuous functions. Real Anal. Exch. **37**(2), 333–351 (2011/2012)
16. G.H. Hardy, *Weierstrass's non-differentiable function*. Trans. Am. Math. Soc. **17**(3), 301–325 (1916)
17. Y. Heurteaux, Weierstrass functions with random phases. Trans. Am. Math. Soc. **355**(8), 3065–3077 (2003)
18. T.Y. Hu, K.-S. Lau, Fractal dimensions and singularities of the Weierstrass type functions. Trans. Am. Math. Soc. **335**(2), 649–665 (1993)
19. P.D. Humke, G. Petruska, The packing dimension of a typical continuous function is 2. Real Anal. Exch. **14**(2), 345–358 (1988/1989)
20. B.R. Hunt, The prevalence of continuous nowhere differentiable functions. Proc. Am. Math. Soc. **122**(3), 711–717 (1994)
21. B.R. Hunt, The Hausdorff dimension of graphs of Weierstrass functions. Proc. Am. Math. Soc. **126**(3), 791–800 (1998)
22. J. Hyde, V. Laschos, L. Olsen, I. Petrykiewicz, A. Shaw, On the box dimensions of graphs of typical continuous functions. J. Math. Anal. Appl. **391**(2), 567–581 (2012)
23. J.-P. Kahane, M. Weiss, G. Weiss, On lacunary power series. Ark. Mat. **5**, 1–26 (1963)

24. J.L. Kaplan, J. Mallet-Paret, J.A. Yorke, The Lyapunov dimension of a nowhere differentiable attracting torus. Ergod. Theory Dyn. Syst. **4**(2), 261–281 (1984)
25. F. Ledrappier, On the dimension of some graphs, in *Symbolic Dynamics and Its Applications*, New Haven, 1991. Contemporary Mathematics, vol. 135 (American Mathematical Society, Providence, 1992), pp. 285–293
26. F. Ledrappier, L.-S. Young, The metric entropy of diffeomorphisms. II. Relations between entropy, exponents and dimension. Ann. Math. (2) **122**(3), 540–574 (1985)
27. B. Mandelbrot, *Fractals: Form, Chance, and Dimension* (W. H. Freeman and Co., San Francisco, 1977)
28. P. Mattila, *Geometry of Sets and Measures in Euclidean Spaces. Fractals and Rectifiability*. Cambridge Studies in Advanced Mathematics, vol. 44 (Cambridge University Press, Cambridge, 1995)
29. R.D. Mauldin, S.C. Williams, On the Hausdorff dimension of some graphs. Trans. Am. Math. Soc. **298**(2), 793–803 (1986)
30. W. Ott, J. Yorke, Prevalence. Bull. Am. Math. Soc. (N.S.) **42**(3), 263–290 (2005)
31. Y. Peres, B. Solomyak, Absolute continuity of Bernoulli convolutions, a simple proof. Math. Res. Lett. **3**(2), 231–239 (1996)
32. Y. Peres, B. Solomyak, Self-similar measures and intersections of Cantor sets. Trans. Am. Math. Soc. **350**(10), 4065–4087 (1998)
33. F. Przytycki, M. Urbański, On the Hausdorff dimension of some fractal sets. Studia Math. **93**(2), 155–186 (1989)
34. F. Rezakhanlou, The packing measure of the graphs and level sets of certain continuous functions. Math. Proc. Camb. Philos. Soc. **104**(2), 347–360 (1988)
35. J. Romanowska, Measure and Hausdorff dimension of randomized Weierstrass-type functions. Nonlinearity **27**(4), 787–801 (2014)
36. F. Roueff, Almost sure Hausdorff dimension of graphs of random wavelet series. J. Fourier Anal. Appl. **9**(3), 237–260 (2003)
37. R. Salem, A. Zygmund, Lacunary power series and Peano curves. Duke Math. J. **12**, 569–578 (1945)
38. P. Shmerkin, On the exceptional set for absolute continuity of Bernoulli convolutions. Geom. Funct. Anal. **24**(3), 946–958 (2014)
39. B. Solomyak, On the random series $\sum \pm\lambda^n$ (an Erdös problem). Ann. Math. (2) **142**(3), 611–625 (1995)
40. B. Solomyak, Measure and dimension for some fractal families. Math. Proc. Camb. Philos. Soc. **124**(3), 531–546 (1998)
41. M. Tsujii, Fat solenoidal attractors. Nonlinearity **14**(5), 1011–1027 (2001)

Tiling $\mathbb{Z}^2$ by a Set of Four Elements

De-Jun Feng and Yang Wang

Abstract A finite subset $\mathcal{D}$ of $\mathbb{Z}^2$ is called a tile of $\mathbb{Z}^2$, if $\mathbb{Z}^2$ can be tiled by disjoint translates of $\mathcal{D}$. In this note, we give a simple characterization of tiles of $\mathbb{Z}^2$ with cardinality 4.

Keywords Tiling • Translation

Mathematics Subject Classification (2000). Primary 05B45; Secondary 52C20

1 Introduction

Let $\mathcal{D}$ be a finite subset of $\mathbb{Z}^d$. We say that $\mathcal{D}$ *tiles* $\mathbb{Z}^d$ or $\mathcal{D}$ *is a tile of* $\mathbb{Z}^d$, if $\mathbb{Z}^d$ can be written as a disjoint union of translates of $\mathcal{D}$, i.e. there exists a subset $\mathcal{C}$ of $\mathbb{Z}^d$ such that each $\mathbf{v} \in \mathbb{Z}^d$ can be expressed uniquely as $\mathbf{x} + \mathbf{y}$ with $\mathbf{x} \in \mathcal{D}$ and $\mathbf{y} \in \mathcal{C}$; in symbols, $\mathbb{Z}^d = \mathcal{D} \oplus \mathcal{C}$. For example, each two-element subset of $\mathbb{Z}$ is a tile of $\mathbb{Z}$, and the set $\{0, 3\}$ tiles $\mathbb{Z}$ in two essentially different ways,

$$\mathbb{Z} = \{0, 3\} \oplus (6\mathbb{Z} \cup (6\mathbb{Z} + 1) \cup (6\mathbb{Z} + 2)) = \{0, 3\} \oplus (2\mathbb{Z}).$$

The first author is supported in part by RGC grants in CUHK (401112, 401013).

D.-J. Feng (✉)
Department of Mathematics, The Chinese University of Hong Kong, Shatin, Hong Kong, China
e-mail: djfeng@math.cuhk.edu.hk

Y. Wang
Department of Mathematics, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China

Department of Mathematics, Michigan State University, East Lansing, MI 48824-1027, USA
e-mail: yangwang@ust.hk

C. Bandt et al. (eds.), *Fractal Geometry and Stochastics V*, Progress in Probability 70, DOI 10.1007/978-3-319-18660-3_6

As an easy fact, $\mathcal{D}$ is a tile of $\mathbb{Z}^d$ if and only if the set $T := \bigcup_{\mathbf{a}\in\mathcal{D}}([0,1]^d + \mathbf{a})$ is a tile of $\mathbb{R}^d$, in the sense that there exists a set $\mathcal{T} \subset \mathbb{R}^d$ such that $\bigcup_{\mathbf{t}\in\mathcal{T}}(T + \mathbf{t}) = \mathbb{R}^d$ and the Lebesgue measure of $(T + \mathbf{t}_1) \cap (T + \mathbf{t}_2)$ is zero for all different $\mathbf{t}_1, \mathbf{t}_2 \in \mathcal{T}$.

One of the fundamental problems in the tiling theory is to characterise the tiles of $\mathbb{Z}^d$ (and generally, $\mathbb{R}^d$) and their tiling structures. The problem is not only interesting to mathematicians, but also to artists, physicists and engineerians (cf. [4]). In mathematics, the classification of tiles of $\mathbb{Z}^d$ is related to the theory of self-similar tiles and self-affine tiles, see [1, 5, 8–13]. It is also related to the Fuglede problem on the tiles and spectral sets which has been received a lot of attention recently [3, 6, 7, 19].

The problem has been studied by many authors in the case when $d = 1$, see [2, 14, 16, 17, 20]. In [14], Newman proved that any tile $\mathcal{D}$ of $\mathbb{Z}$ tiles $\mathbb{Z}$ periodically, that is, if $\mathcal{D} \oplus \mathcal{C} = \mathbb{Z}$ then $\mathcal{C} = \mathcal{C} + v$ for some $v \in \mathbb{Z}$. Furthermore, Newman gave a simple characterization of tiles $\mathcal{D}$ of $\mathbb{Z}$ in the case that the cardinality of $\mathcal{D}$ is a prime power. To state his result, let $\mathcal{D} = \{a_1, \ldots, a_k\}$ with $k = \#\mathcal{D} = p^\alpha$, where p is a prime and $\alpha \in \mathbb{N}$. For any pair a_i, a_j, $i \neq j$, let $t_{i,j}$ be the highest power of p which divides $a_i - a_j$. Newman showed that $\mathcal{D}$ tiles $\mathbb{Z}$ if and only if there are at most α distinct $t_{i,j}$. Applying this criterion to the special case that $\#\mathcal{D} = 3$, we see that if $\mathcal{D}$ tiles $\mathbb{Z}$ and if not all the difference of the numbers in $\mathcal{D}$ are divisible by 3, then no difference could be a multiple of 3. Due to this, $\{0, 1, 3\}$ and $\{0, 1, 4\}$ do not tile $\mathbb{Z}$. Later, Coven and Meyerowitz [2] did the work in the case that the cardinality of $\mathcal{D}$ has at most two prime factors. For arbitrary cardinality, the problem is still open.

The case $d \geq 2$ is much harder to study. One of the difficulties comes from the possibility that a tile of $\mathbb{Z}^d$ with $d \geq 2$, might admit a non-periodic tiling of $\mathbb{Z}^d$. In [8], Lagarias and Wang conjectured that any tile $\mathcal{D}$ of $\mathbb{Z}^d$ can tile $\mathbb{Z}^d$ periodically when $d \geq 2$. However, this conjecture is still wide open. Using a group-theoretic approach, Szegedy [18] proved the conjecture in the particular case that $d = 2$ and $\#\mathcal{D}$ is a prime or equals 4. Furthermore, he formulated two algorithms to check whether a given $\mathcal{D} \subset \mathbb{Z}^2$ can tile $\mathbb{Z}^2$, in the two different situations: (i) $\mathcal{D}$ is a prime; (ii) $\mathcal{D} = 4$. His algorithm in dealing with the prime case was re-discovered by Rao and Xue in [15] under an additional assumption of periodicity.

It is a simple fact that any two-element subset of $\mathbb{Z}^2$ tiles $\mathbb{Z}^2$. For the convenience of the reader, we present here the simple classification for $\mathcal{D} = 3$: each three-element subset of $\mathbb{Z}^2$ that is not contained in a straight line always tiles $\mathbb{Z}^2$; and if the subset is contained in a straight line, then it tiles $\mathbb{Z}^2$ if and only if it tiles that straight line. The second part of the classification is easy to see, and we may use Newman's criterion to check whether the set tiles the straight line. To see the first part, we just take an example, say $\mathcal{D} = \{(0,0), (m,0), (0,1)\}$, to illustrate the idea. Choose an integer n, different from $0, m$, such that $\{0, m, n\}$ tiles $\mathbb{Z}$ (the existence of such n follows from Newman's criterion). Take $\mathcal{T} \subset \mathbb{Z}$ so that $\mathbb{Z} = \{0, m, n\} \oplus \mathcal{T}$. Set

$$\mathcal{C} = \{(a - kn, k) : a \in \mathcal{T}, k \in \mathbb{Z}\}.$$

Then it is easy to check that $\mathbb{Z}^2 = \mathcal{D} \oplus \mathcal{C}$.

In this article, we give the following characterization of tiles of $\mathbb{Z}^2$ with cardinality 4, which looks much more intuitive and simpler than Szegedy's algorithm.

Theorem 1.1 *Let $\mathcal{D} \subset \mathbb{Z}^2$ with $\#\mathcal{D} = 4$. Then $\mathcal{D}$ is not a tile of $\mathbb{Z}^2$ if and only if $\mathcal{D}$ is of the form $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ in which $\mathbf{v}_2 - \mathbf{v}_1 = \frac{p}{q}(\mathbf{v}_4 - \mathbf{v}_3)$ for some $p \in 2\mathbb{Z} \setminus \{0\}$ and $q \in 2\mathbb{Z} + 1$.*

According to the above theorem, $\mathcal{D} \subset \mathbb{Z}^2$ with cardinality 4 is not a tile of $\mathbb{Z}^2$ if and only if one of the following two scenarios occurs: (1) $\mathcal{D}$ is the set of vertices of a bilateral which contains two parallel edges with length ratio in $2\mathbb{Z}/(2\mathbb{Z} + 1)$; (2) $\mathcal{D}$ is contained in a straight line and satisfies the ratio condition in the theorem. See Fig. 1 for their geometric pictures.

We remark that using Theorem 1.1 and the involved proofs, we can manage to prove that Fuglede's conjecture is true for the union of four integral unit squares in $\mathbb{R}^2$, that is, if $\Omega = \bigcup_{\mathbf{v} \in \mathcal{D}} [0, 1]^2 + \mathbf{v}$ for some $\mathcal{D} \subset \mathbb{Z}^2$ with $\#\mathcal{D} = 4$, then Ω tiles $\mathbb{R}^2$ by translation if and only if $L^2(\Omega)$ has an orthogonal basis consisting of complex exponentials. Details will be given somewhere else.

2 A Standard Form for $\mathcal{D}$

First suppose that $\mathcal{D}$ is contained in a straight line in $\mathbb{R}^2$. Without loss of generality, we may just assume that $\mathcal{D}$ is on the x-axis, i.e. $\mathcal{D} \subseteq \mathbb{Z}$. In this case, it is clear that $\mathcal{D}$ tiles $\mathbb{Z}^2$ if and only if $\mathcal{D}$ tiles $\mathbb{Z}$. Write $\mathcal{D} = \{v_1, v_2, v_3, v_4\}$. For $1 \le i, j \le 4$ with $i \ne j$, let $2^{e_{i,j}}$ be the highest power of 2 which divides $v_i - v_j$, where $1 \le i, j \le 4$ and $i \ne j$. Newman [14] proved that $\mathcal{D}$ tiles $\mathbb{Z}$ if and only if there are at most 2 distinct $e_{i,j}$. We will show that this criterion implies Theorem 1.1. For convenience, we say that $\mathcal{D}$ has *property (I)* if there are at least 3 distinct $e_{i,j}$, whilst we say that $\mathcal{D}$ has *property (II)* if there is a partition $\{v_1', v_2'\} \cup \{v_3', v_4'\}$ of $\mathcal{D}$ such that $v_2' - v_1' = \frac{p}{q}(v_4' - v_3')$ for some $p \in 2\mathbb{Z} \setminus \{0\}$ and $q \in 2\mathbb{Z} + 1$. We claim that these two properties are equivalent, from which Theorem 1.1 follows.

To see the claim, we notice that these two properties keep unchanged if we replace $\mathcal{D}$ by an affine copy of $\mathcal{D}$, provided that the latter is contained in $\mathbb{Z}$. Hence we may assume, without loss of generality, that $\mathcal{D}$ contains 0 and at least an odd integer.

Property (II) $\Longrightarrow$ property (I). Suppose that D has a partition $\{v_1, v_2\} \cup \{v_3, v_4\}$ such that

$$v_2 - v_1 = \frac{p}{q}(v_4 - v_3)$$

with $p \in 2\mathbb{Z}$ and $q \in 2\mathbb{Z} + 1$. Equivalently, $e_{1,2} > e_{3,4}$. Suppose that property (I) does not hold, i.e., $e_{i,j} \in \{e_{1,2}, e_{3,4}\}$ for all $i \ne j$. We derive a contradiction as follows. Notice that $e_{1,3} \ne e_{1,2}$ (since $e_{1,3} = e_{1,2}$ implies that $e_{2,3} > e_{1,2}$), and also

$e_{1,4} \neq e_{1,2}$. Hence $e_{1,3} = e_{3,4}$ and $e_{1,4} = e_{3,4}$. But this is impossible since $e_{1,3} = e_{1,4}$ implies that $e_{3,4} > e_{1,3}$.

Property (I) $\Longrightarrow$ property (II). Suppose that property (I) holds. When $\mathcal{D}$ contains 0 and three odd numbers, or $\mathcal{D}$ contains 0, one odd number and two non-zero even numbers, it is easy to see that property (II) holds. The remaining case is that $\mathcal{D} = \{0, a, b, c\}$, where a is a non-zero even number, and b, c are two odd numbers. The assumption that there are at least 3 distinct $e_{i,j}$ then implies that the highest power of 2 dividing $b - c$ is different from that dividing $a - 0 = a$, from which property (I) follows. This completes the proof of the claim.

In what follows, we consider the case that $\mathcal{D}$ is not contained in a straight line. We introduce a standard form for our set $\mathcal{D}$ and prove the following statement instead of Theorem 1.1 (here and afterwards, we will use column vectors to express elements in $\mathbb{Z}^2$).

Theorem 2.1 *Let $\mathcal{D}$ be a subset of $\mathbb{Z}^2$ with cardinality 4. Assume that $\mathcal{D}$ is not contained in a line, and furthermore $\mathbf{0} \in \mathcal{D}$. Then a sufficient and necessary condition for $\mathcal{D}$ not to tile $\mathbb{Z}^2$ is that there exists a 2×2 matrix G so that*

$$G\mathcal{D} = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ p/q \end{pmatrix} \right\}$$

with $p, q \in \mathbb{Z} \setminus \{0\}$ and $p + q \in 2\mathbb{Z} + 1$;

The equivalence of the above theorem with Theorem 1.1 comes from the following simple fact: if $\mathcal{D}$ (which contains the origin) is the set of vertices of a bilateral which contains two parallel edges with length ratio in $2\mathbb{Z}/(2\mathbb{Z} + 1)$, then there exists an invertible linear transformation G such that $G\mathcal{D}$ has the standard form given in Theorem 2.1. To see this, let $\mathcal{D} = \{A, B, C, D\}$ as shown in Fig. 1, where the edge AB is parallel to CD, and assume $A = \mathbf{0}$. Let G be the linear transformation so that $GD = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $GB = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Then GC has the desired form $\begin{pmatrix} 1 \\ p/q \end{pmatrix}$ with $p, q \in \mathbb{Z} \backslash \{0\}$, and $p + q \in 2\mathbb{Z} + 1$.

To prove Theorem 2.1, we need the following.

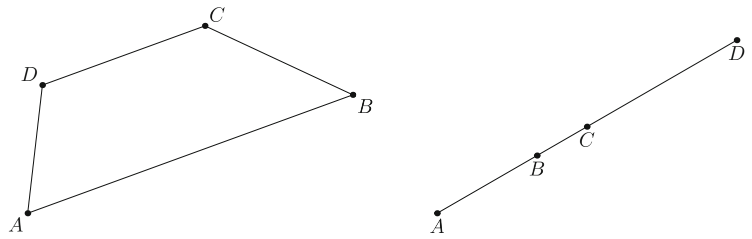


Fig. 1 Two general cases when $\mathcal{D}$ can not tile $\mathbb{Z}^2$

Proposition 2.2 *Let $\mathcal{A}$ be a finite subset of $\mathbb{Z}^d$. Then the following statements are equivalent:*

(i) *There exists $\mathcal{B} \subset \mathbb{Z}^d$ such that $\mathcal{A} \oplus \mathcal{B} = \mathbb{Z}^d$.*
(ii) *There exist a non-singular $d \times d$ matrix G with rational entries and $\mathcal{D} \subset \mathbb{Q}^d$ such that $(G\mathcal{A}) \oplus \mathcal{D}$ is a lattice in $\mathbb{R}^d$.*

Proof The direction (i) $\Longrightarrow$ (ii) is clear. Now we show the opposite direction. Assume that $(G\mathcal{A}) \oplus \mathcal{D}$ is a lattice in $\mathbb{R}^d$, i.e. $(G\mathcal{A}) \oplus \mathcal{D} = H\mathbb{Z}^2$ for some $d \times d$ matrix H. Clearly H is rational. We may assume that H is non-singular (otherwise there exists $\mathcal{C} \subset \mathbb{Q}^d$ so that $H\mathbb{Z}^d \oplus \mathcal{C} = \tilde{H}\mathbb{Z}^d$ for some non-singular $d \times d$ rational matrix $\tilde{H}$ and $(G\mathcal{A}) \oplus \mathcal{D} \oplus \mathcal{C} = \tilde{H}\mathbb{Z}^2$). Then $(H^{-1}G\mathcal{A}) \oplus (H^{-1}\mathcal{D}) = \mathbb{Z}^d$. Choose an integer p so that $E := pH^{-1}G$ is an integral matrix. Note that $(E\mathcal{A}) \oplus (pH^{-1}\mathcal{D}) = p\mathbb{Z}^d$ and $E\mathcal{A} \subset \mathbb{Z}^d$. It follows that $\Lambda := (pH^{-1}\mathcal{D}) \subset \mathbb{Z}^d$. Since p is an integer, there exists a finite set $V \subset \mathbb{Z}^d$ so that $(p\mathbb{Z}^d) \oplus V = \mathbb{Z}^d$. Therefore $(E\mathcal{A}) \oplus \Lambda \oplus V = \mathbb{Z}^d$. Note that $\mathbb{Z}^d = (E\mathbb{Z}^d) \oplus U$ for some finite set $U \subset \mathbb{Z}^d$ with $\mathbf{0} \in U$. We have $(E\mathcal{A}) \oplus \Lambda \oplus V = (E\mathbb{Z}^d) \oplus U$. Letting $\tilde{\Lambda} = (E\mathbb{Z}^d) \cap (\Lambda \oplus V)$, we obtain $(E\mathcal{A}) \oplus \tilde{\Lambda} = E\mathbb{Z}^d$. This implies $\mathcal{A} \oplus (E^{-1}\tilde{\Lambda}) = \mathbb{Z}^d$. □

As a direct corollary of the above proposition, we have

Corollary 2.3 *Let $\mathcal{A}$ and $\mathcal{B}$ be two finite subsets of $\mathbb{Z}^d$. If $\mathcal{A} = G\mathcal{B}$ for some non-singular $d \times d$ rational matrix G, then $\mathcal{A}$ tiles $\mathbb{Z}^d$ if and only if $\mathcal{B}$ tiles $\mathbb{Z}^d$.*

3 Construction of Tilings

Lemma 3.1 *Let*

$$\mathcal{C} = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} p_1 \\ \frac{1}{2} + p_2 \end{pmatrix}, \begin{pmatrix} \frac{1}{2} + p_3 \\ t + p_4 \end{pmatrix}, \begin{pmatrix} \frac{1}{2} + p_5 \\ \frac{1}{2} + t + p_6 \end{pmatrix} \right\} \tag{3.1}$$

for some $t \in \mathbb{Q}$ and $p_j \in \mathbb{Z}$ ($j = 1, \ldots, 6$). Then there exists $\mathcal{B} \subset \mathbb{Q}^2$ such that $\mathcal{C} \oplus \mathcal{B}$ is a lattice in $\mathbb{R}^2$.

Proof Write $t = \frac{m}{n}$, where $m \in \mathbb{Z}$, $n \in \mathbb{N}$ and $\gcd(m, n) = 1$. Define

$$\mathcal{B} = \left\{ \begin{pmatrix} u \\ v + \frac{j}{2n} \end{pmatrix} : \ u, v, j \in \mathbb{Z}, \ 0 \le j \le n - 1 \right\}.$$

Then by direct calculation, we conclude

$$\mathcal{C} \oplus \mathcal{B} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2n} \end{pmatrix} \mathbb{Z}^2.$$

This finishes the proof. □

Corollary 3.2 *Let $\mathcal{A}$ be a subset of $\mathbb{Z}^2$ of cardinality* 4. *Assume* $\mathbf{0} \in \mathcal{A}$. *Then $\mathcal{A}$ can tile $\mathbb{Z}^2$ if $\mathcal{A} = G\mathcal{C}$ for some 2×2 non-singular rational matrix G and some $\mathcal{C} \subset \mathbb{R}^2$ of the form as in* (3.1).

Proof By Lemma 3.1, there exists $B \subset \mathbb{Q}^2$ such that $\mathcal{C} \oplus \mathcal{B}$ is a lattice in $\mathbb{R}^2$. That is, $(G^{-1}\mathcal{A}) \oplus \mathcal{B}$ is a lattice. Hence by Proposition 2.2, $\mathcal{A}$ can tile $\mathbb{Z}^2$. □

Lemma 3.3 *Let $u, v \in \mathbb{Q}$. Then at least one of the following three equations has a solution $(x, y, z) \in \mathbb{Z}^3$:*

(i) $(\frac{1}{2} + x)u + (\frac{1}{2} + y)v = z.$
(ii) $xu + (\frac{1}{2} + y)v = \frac{1}{2} + z.$
(iii) $(\frac{1}{2} + x)u + yv = \frac{1}{2} + z.$

Proof We first show that if one of the equations has an integral solution (x_0, y_0, z_0) for u, v and p, q, p', q' are integers then there is an integral solution (x_1, y_1, z_1) for $\tilde{u} = u(2p + 1)/(2q + 1)$ and $\tilde{v} = v(2p' + 1)/(2q' + 1)$. For brevity, we give the proof only for equation (i). Take x_1, y_1 so that

$$(\frac{1}{2} + x_1) = (\frac{1}{2} + x_0)(2q + 1)(2p' + 1), \quad (\frac{1}{2} + y_1) = (\frac{1}{2} + y_0)(2q' + 1)(2p + 1).$$

Then $x_1, y_1 \in \mathbb{Z}$. Set $z_1 = z_0(2p + 1)(2p' + 1)$. Then (x_1, y_1, z_1) satisfies

$$(\frac{1}{2} + x_1)\tilde{u} + (\frac{1}{2} + y_1)\tilde{v} = z_1.$$

This proves the claim.

Hence to prove the lemma, we can assume without loss of generality that $u = 2^m$ and $v = 2^n$ for $m, n \in \mathbb{Z}$. This case is not difficult to check. □

For convenience, set

$$\mathbb{Q}_1 = \{p/q : \ p, q \text{ are odd integers}\}. \tag{3.2}$$

Proposition 3.4 *Let*

$$\mathcal{D} = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} u \\ v \end{pmatrix} \right\}$$

where $u, v \in \mathbb{Q}$. Then there exists a non-singular 2×2 rational matrix G such that $G\mathcal{D}$ has the same form as $\mathcal{C}$ in Lemma 3.1 if the pair (u, v) does not satisfy anyone of the following conditions:

(i) $u = 1$ *and* $v \notin \mathbb{Q}_1$.
(ii) $v = 1$ *and* $u \notin \mathbb{Q}_1$.
(iii) $u = -v$ *and* $u \notin \mathbb{Q}_1$.

Proof We will prove the existence of G in each of the following scenarios:

1. $u = 1$ and $v \in \mathbb{Q}_1$.
2. $v = 1$ and $u \in \mathbb{Q}_1$.
3. $u = -v$ and $u \in \mathbb{Q}_1$.
4. $u \neq 1$, $v \neq 1$ and $u \neq -v$.

In scenario (1), let $v = \frac{2q+1}{2p+1}$, where $p, q \in \mathbb{Z}$. We may take

$$G = \begin{pmatrix} 0 & p + \frac{1}{2} \\ \frac{1}{2} & 2p + 1 \end{pmatrix}.$$

Then $G\mathcal{D} = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ \frac{1}{2} \end{pmatrix}, \begin{pmatrix} \frac{1}{2} + p \\ 2p + 1 \end{pmatrix}, \begin{pmatrix} \frac{1}{2} + q \\ \frac{1}{2} + 2q + 1 \end{pmatrix} \right\}$.

In scenario (2), let $u = \frac{2q+1}{2p+1}$, where $p, q \in \mathbb{Z}$. We may take

$$G = \begin{pmatrix} p + \frac{1}{2} & 0 \\ 2p + 1 & \frac{1}{2} \end{pmatrix}.$$

Then $G\mathcal{D}$ has the same expression as that in scenario (1).

In scenario (3), let $u = \frac{2q+1}{2p+1}$, where $p, q \in \mathbb{Z}$. We may take

$$G = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 2p + 1 & p + \frac{1}{2} \end{pmatrix}.$$

Then $G\mathcal{D} = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{1}{2} \\ 2p + 1 \end{pmatrix}, \begin{pmatrix} \frac{1}{2} \\ p + \frac{1}{2} \end{pmatrix}, \begin{pmatrix} 0 \\ q + \frac{1}{2} \end{pmatrix} \right\}$.

Now let us turn to the scenario (4). By Lemma 3.3, one of the following equations has an integral solution (x, y, z):

(e1) $(\frac{1}{2} + x)u + (\frac{1}{2} + y)v = z$.
(e2) $xu + (\frac{1}{2} + y)v = \frac{1}{2} + z$.
(e3) $(\frac{1}{2} + x)u + yv = \frac{1}{2} + z$.

Assume at first that (e1) has an integral solution (x, y, z). Since $u \neq -v$, there exists $t \in \mathbb{Q}$ such that

$$(t + x)u + (\frac{1}{2} + t + y)v = \frac{1}{2} + z.$$

Take $G = \begin{pmatrix} \frac{1}{2}+x & \frac{1}{2}+y \\ t+x & \frac{1}{2}+t+y \end{pmatrix}$. Then

$$G\mathcal{D} = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{1}{2}+x \\ t+x \end{pmatrix}, \begin{pmatrix} \frac{1}{2}+y \\ \frac{1}{2}+t+y \end{pmatrix}, \begin{pmatrix} z \\ \frac{1}{2}+z \end{pmatrix} \right\}.$$

Now we assume (e2) has an integral solution (x, y, z). Since $v \neq 1$, there exists $t \in \mathbb{Q}$ so that $\frac{1}{2}u + tv = \frac{1}{2} + t$. Take $G = \begin{pmatrix} x & \frac{1}{2}+y \\ \frac{1}{2} & t \end{pmatrix}$. Then

$$G\mathcal{D} = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} x \\ \frac{1}{2} \end{pmatrix}, \begin{pmatrix} \frac{1}{2}+y \\ t \end{pmatrix}, \begin{pmatrix} \frac{1}{2}+z \\ \frac{1}{2}+t \end{pmatrix} \right\}.$$

If (e3) has an integral solution, we may construct G in a similar way as above. □

4 Nonexistence of Tilings

Proposition 4.1 *Let*

$$\mathcal{D} = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} u \\ v \end{pmatrix} \right\}$$

where $u, v \in \mathbb{Q}$. Then there exists no $\mathcal{C}$ such that $\mathcal{D} \oplus \mathcal{C}$ is a lattice if u, v does satisfy one of the following conditions:

(i) $u = 1$ *and* $v \notin \mathbb{Q}_1$;
(ii) $v = 1$ *and* $u \notin \mathbb{Q}_1$;
(iii) $u = -v$ *and* $u \notin \mathbb{Q}_1$;

where $\mathbb{Q}_1$ is defined by (3.2).

Proof Without loss of generality we may only consider case (ii), since the sets $\mathcal{D}$ in cases (i) and (iii) differ from that in (ii) only by an affine map.

Assume $u = \frac{p}{q}$ with $p \in \mathbb{Z}$, $q \in \mathbb{N}$ and $p + q \in 2\mathbb{Z} + 1$. Take $G = \begin{pmatrix} q & 0 \\ 0 & 1 \end{pmatrix}$. Then

$$G\mathcal{D} = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} q \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} p \\ 1 \end{pmatrix} \right\}$$

By Proposition 2.2, we only need to prove that $G\mathcal{D}$ can not tile $\mathbb{Z}^2$.

Assume on the contrary that $G\mathcal{D}$ can tile $\mathbb{Z}^2$, i.e., $(G\mathcal{D}) \oplus \Lambda = \mathbb{Z}^2$. Then any $\mathbf{x} \in \mathbb{Z}^2$ can be uniquely written as $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$ with $\mathbf{x}_1 \in G\mathcal{D}$ and $\mathbf{x}_2 \in \Lambda$. Define

$\phi : \mathbb{Z}^2 \to G\mathcal{D}$ by $\mathbf{x} \mapsto \mathbf{x}_1$. Let $\{a_n\}_{n\in\mathbb{Z}}$ be the sequence defined by

$$a_n = \begin{cases} 1 & \text{if } \phi(n,0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ 2 & \text{if } \phi(n,0) = \begin{pmatrix} q \\ 0 \end{pmatrix} \\ 3 & \text{if } \phi(n,0) = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ 4 & \text{if } \phi(n,0) = \begin{pmatrix} p \\ 1 \end{pmatrix} \end{cases}.$$

We have the following observations:

(a) For any $n \in \mathbb{Z}$, $a_{n+p} \neq a_n$ and $a_{n+q} \neq a_n$.
(b) If $a_n = 1$ then $a_{n+q} = 2$. If $a_n = 2$ then $a_{n-q} = 1$. If $a_n = 3$ then $a_{n+p} = 4$. If $a_n = 4$ then $a_{n-p} = 3$.

Let us first prove (a). From $(G\mathcal{D}) \oplus \Lambda = \mathbb{Z}^2$ we obtain $(G\mathcal{D} - G\mathcal{D}) \cap (\Lambda - \Lambda) = \{\mathbf{0}\}$. Since $\begin{pmatrix} p \\ 0 \end{pmatrix}, \begin{pmatrix} q \\ 0 \end{pmatrix} \in G\mathcal{D} - G\mathcal{D}$, we have $\begin{pmatrix} p \\ 0 \end{pmatrix}, \begin{pmatrix} q \\ 0 \end{pmatrix} \notin \Lambda - \Lambda$. Now assume (a) is not true. Without loss of generality we assume $a_{n+p} = a_n$ for some n. Then

$$\begin{pmatrix} n \\ 0 \end{pmatrix} = \mathbf{y} + \boldsymbol{\lambda}_1, \qquad \begin{pmatrix} n+p \\ 0 \end{pmatrix} = \mathbf{y} + \boldsymbol{\lambda}_2$$

for some $\mathbf{y} \in G\mathcal{D}$ and $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2 \in \Lambda$. It implies that $\begin{pmatrix} p \\ 0 \end{pmatrix} = \boldsymbol{\lambda}_2 - \boldsymbol{\lambda}_1 \in \Lambda - \Lambda$, which leads to a contradiction. This proves (a). To prove (b) without loss of generality we prove that $a_{n+q} = 2$ when $a_n = 1$. Since $a_n = 1$, we have $\begin{pmatrix} n \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \boldsymbol{\lambda}$ for some $\boldsymbol{\lambda} \in \Lambda$. Therefore $\begin{pmatrix} n+q \\ 0 \end{pmatrix} = \begin{pmatrix} q \\ 0 \end{pmatrix} + \boldsymbol{\lambda}$, which implies $a_{n+q} = 2$. This finishes the proof of (b).

According to (a) and (b), we have the following claims:

(c1) Assume $p > 0$. If $a_n \in \{1,3\}$, then $a_{n+p+q} \in \{1,3\}$.
(c2) Assume $p < 0$. If $a_n \in \{1,4\}$, then $a_{n-p+q} \in \{2,3\}$.

Without loss of generality we only prove (c1). First assume $a_n = 1$. Then by (b) we have $a_{n+q} = 2$. Thus by (a) we have $a_{n+p+q} \neq 2$. In the same time by (b) we have $a_{n+p+q} \neq 4$ since otherwise $a_{n+q} = 3$. Therefore we always have $a_{n+p+q} \in \{1,3\}$ when $a_n = 1$. Using an essentially identical argument, we can prove that $a_{n+p+q} \in \{1,3\}$ when $a_n = 3$. This finishes the proof of (c1).

Now assume $p > 0$. Then (c1) implies that the set $\{0, 1, \ldots, p + q - 1\}$ can be partitioned into two sets A and B such that there exists a large $N \in \mathbb{N}$ so that for $n > N$, $a_n \in \{1, 3\}$ if $n (\text{mod } p + q) \in A$, and $a_n \in \{2, 4\}$ if $n (\text{mod } p + q) \in B$. That means the density of those n with $a_n \in \{1, 3\}$ in $\mathbb{Z} \cap [N, \infty)$ is $\#A/(p+q)$, and the density of the rest is $\#B/(p + q)$. Since $p + q \in 2\mathbb{Z} + 1$, these two densities are different. However from (b), these two densities must be the same. This leads to a contradiction.

Similarly we can derive a contradiction in the case when $p < 0$. This finishes the proof of the proposition. □

5 Proof of Theorem 2.1

Proof Since $\mathcal{D}$ is not contained in a straight line, there exists a non-singular rational 2×2 matrix A so that

$$A\mathcal{D} = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} u \\ v \end{pmatrix} \right\}$$

with $u, v \in \mathbb{Q}$. Assume $\mathcal{D}$ can not tile $\mathbb{Z}^2$. Then by Proposition 2.2, there is no non-singular rational matrix G and $\mathcal{C} \subset \mathbb{Q}^2$ such that $G\mathcal{D} \oplus \mathcal{C}$ is a lattice. Therefore by Proposition 3.4 and Lemma 3.1, u, v do satisfy one of the following conditions:

(i) $u = 1$ and $v \notin \mathbb{Q}_1$;
(ii) $v = 1$ and $u \notin \mathbb{Q}_1$;
(iii) $u = -v$ and $u \notin \mathbb{Q}_1$;

where $\mathbb{Q}_1$ is defined by (3.2). By taking B to be

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & 1 \\ 0 & -1/u \end{pmatrix}$$

respectively, in the above 3 cases, we see that

$$BA\mathcal{D} = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ p/q \end{pmatrix} \right\}$$

for some $p, q \in \mathbb{Z} \setminus \{0\}$ with $p + q \in 2\mathbb{Z} + 1$. This proves the necessity. The sufficiency is implied by Proposition 4.1. □

References

1. C. Bandt, Self-similar sets. V. Integer matrices and fractal tilings of $\mathbb{R}^n$. Proc. Am. Math. Soc. **112**(2), 549–562 (1991)
2. E. Coven, A. Meyerowitz, Tiling the integers with translates of one finite set. J. Algebra **212**, 161–174 (1999)
3. B. Fuglede, Commuting self-adjoint partial differential operators and a group theoretic problem. J. Funct. Anal. **16**, 101–121 (1974)
4. B. Grünbaum, G.C. Shephard, *Tilings and Patterns. An Introduction*. A Series of Books in the Mathematical Sciences (W. H. Freeman, New York, 1989)
5. R. Kenyon, Self-replicating tilings, in *Symbolic Dynamics and Its Applications*, ed. by P. Walters (American Mathematical Society, Providence, 1992), pp. 239–264
6. M. Kolountzakis, The study of translation tiling with Fourier analysis, in *Fourier Analysis and Convexity*, ed. by L. Brandolini et al. Applied and Numerical Harmonic Analysis (Birkhäuser, Boston, 2004), pp. 131–187
7. M. Kolountzakis, Periodicity of the spectrum of a finite union of intervals. J. Fourier Anal. Appl. **18**(1), 21–26 (2012)
8. J. Lagarias, Y. Wang, Tiling the line with translates of one tile. Invent. Math. **124**, 341–365 (1996)
9. J. Lagarias, Y. Wang, Integral self-affine tiles in $\mathbb{R}^n$. I. Standard and nonstandard digit sets. J. Lond. Math. Soc. **54**(1), 161–179 (1996)
10. J. Lagarias, Y. Wang, Self-affine tiles in $\mathbb{R}^n$. Adv. Math. **121**(1), 21–49 (1996)
11. J. Lagarias, Y. Wang, Integral self-affine tiles in $\mathbb{R}^n$. II. Lattice tilings. J. Fourier Anal. Appl. **3**(1), 83–102 (1997)
12. C.K. Lai, K.S. Lau, H. Rao, Spectral structure of digit sets of self-similar tiles on $\mathbb{R}^1$. Trans. Am. Math. Soc. **365**(7), 3831–3850 (2013)
13. K.S. Lau, H. Rao, On one-dimensional self-similar tilings and the *pq*-tilings. Trans. Am. Math. Soc. **355**, 1401–1414 (2003)
14. D.J. Newman, Tesselation of integers. J. Number Theory **9**, 107–111 (1977)
15. H. Rao, Y.M. Xue, Tiling $\mathbb{Z}^2$ with translations of one set. Discret. Math. Theor. Comput. Sci. **8**, 129–140 (2006)
16. A.D. Sands, On Keller's conjecture for certain cyclic groups. Proc. Edinb. Math. Soc. **22**, 17–21 (1979)
17. S. Szabó, A type of factorization of finite abelian groups. Discret. Math. **54**, 121–124 (1985)
18. M. Szegedy, Algorithms to tile the infinite grid with finite clusters, in *Proceedings of the 39th Annual Symposium on the Foundations of Computer Science*, Palo Alto, 1998, pp. 137–145
19. T. Tao, Fuglede's conjecture is false in 5 or higher dimensions. Math. Res. Lett. **11**, 251–258 (2004)
20. R. Tijdeman, Decomposition of the integers as a direct sum of two subsets, in *Number Theory*, Paris, 1992–1993. London Mathematical Society Lecture Note Series, vol. 215 (Cambridge University Press, Cambridge, 1995), pp. 261–276

Some Recent Developments in Quantization of Fractal Measures

Marc Kesseböhmer and Sanguo Zhu

Abstract We give an overview on the quantization problem for fractal measures, including some related results and methods which have been developed in the last decades. Based on the work of Graf and Luschgy, we propose a three-step procedure to estimate the quantization errors. We survey some recent progress, which makes use of this procedure, including the quantization for self-affine measures, Markov-type measures on graph-directed fractals, and product measures on multiscale Moran sets. Several open problems are mentioned.

Keywords Quantization dimension • Quantization coefficient • Bedford-McMullen carpets • Self-affine measures • Markov measures • Moran measures

Mathematics Subject Classification (2000). Primary 28A75, Secondary 28A80, 94A15

1 Introduction

The quantization problem for probability measures originated in information theory and certain areas of engineering technology such as image compression and data processing. In the past decades, this problem has been rigorously studied by mathematicians and the field of quantization theory emerged. Recently, this theory

---

S. Zhu was supported by China Scholarship Council No. 201308320049

M. Kesseböhmer (✉)
Fachbereich 3 – Mathematik und Informatik, Universität Bremen, Bibliothekstr. 1, Bremen 28359, Germany
e-mail: mhk@math.uni-bremen.de

S. Zhu
School of Mathematics and Physics, Jiangsu University of Technology, Changzhou 213001, China
e-mail: sgzhu@jsut.edu.cn

C. Bandt et al. (eds.), *Fractal Geometry and Stochastics V*, Progress in Probability 70, DOI 10.1007/978-3-319-18660-3_7

has also been found to have promising applications in numerical integrations and mathematical finance (see e.g. [22–24]). Mathematically we are concerned with the asymptotics of the errors in the approximation of a given probability measure with finitely supported probability measures in the sense of L_r-metrics. More precisely, for every $n \in \mathbb{N}$, we set $\mathcal{D}_n := \{\alpha \subset \mathbb{R}^q : 1 \le \mathrm{card}(\alpha) \le n\}$. Let μ be a Borel probability measure on $\mathbb{R}^q$, $q \in \mathbb{N}$, and let $r \in [0, \infty)$. The n-th quantization error for μ of order r is given by [6]

$$e_{n,r}(\mu) := \begin{cases} \inf_{\alpha \in \mathcal{D}_n} \left(\int d(x,\alpha)^r \, \mathrm{d}\mu(x)\right)^{1/r}, & r > 0, \\ \inf_{\alpha \in \mathcal{D}_n} \exp \int \log d(x,\alpha) \, \mathrm{d}\mu(x), & r = 0. \end{cases} \tag{1.1}$$

According to [6], $e_{n,r}(\mu)$ equals the error with respect to the L_r-metrics in the approximation of μ with discrete probability measures supported on at most n points. See [6, 13] for various equivalent definitions for the quantization error. In the following we will focus on the L_r-quantization problem with $r > 0$. For the quantization with respect to the geometric mean error, we refer to [8] for rigorous foundations and [37, 41, 42, 44] for more related results.

The upper and lower quantization dimension for μ of order r, as defined below, characterize the asymptotic quantization error in a natural manner:

$$\overline{D}_r(\mu) := \limsup_{n\to\infty} \frac{\log n}{-\log e_{n,r}(\mu)}, \quad \underline{D}_r(\mu) := \liminf_{n\to\infty} \frac{\log n}{-\log e_{n,r}(\mu)}.$$

If $\overline{D}_r(\mu) = \underline{D}_r(\mu)$, we call the common value the quantization dimension of μ of order r and denote it by $D_r(\mu)$. To obtain more accurate information about the asymptotic quantization error, we define the s-dimensional upper and lower quantization coefficient (cf. [6, 26]):

$$\overline{Q}_r^s(\mu) := \limsup_{n\to\infty} n^{1/s} e_{n,r}(\mu), \quad \underline{Q}_r^s(\mu) := \liminf_{n\to\infty} n^{1/s} e_{n,r}(\mu), \quad s > 0.$$

By [6, 26], the upper (lower) quantization dimension is exactly the critical point at which the upper (lower) quantization coefficient jumps from zero to infinity.

The following theorem by Zador is a classical result on quantization of absolutely continuous measures. It was first proposed by Zador [32] and then generalized by Bucklew and Wise [2]; we refer to [6, Theorem 6.2] for a rigorous proof.

Theorem 1.1 ([6]) *Let μ be absolutely continuous Borel probability measure on $\mathbb{R}^q$ with density h with respect to the q-dimensional Lebesgue measure λ^q. Assume that for some $\delta > 0$, we have $\int |x|^{r+\delta} \, \mathrm{d}\mu\,(x) < \infty$. Then for all $r > 0$ we have*

$$\underline{Q}_r^q(\mu) = \overline{Q}_r^q(\mu) = C(r,q)\left(\int h^{\frac{q}{q+r}}(x) \, \mathrm{d}\lambda^q\,(x)\right)^{\frac{q+r}{q}},$$

where $C(r,q)$ is a constant independent of μ.

While engineers are mainly dealing with absolutely continuous distributions, the quantization problem is significant for all Borel probability measures satisfying the moment condition $\int |x|^r \mathrm{d}\mu(x) < \infty$. For later use we define the subset of Borel probabilities $\mathcal{M}_r := \left\{\mu \colon \mu(\mathbb{R}) = 1, \int |x|^r \mathrm{d}\mu(x) < \infty\right\}$ and let $\mathcal{M}_\infty$ denote the set of Borel probability measures with compact support. This condition ensures that the set of n-optimal sets of order r denoted by $C_{n,r}(\mu)$ is non-empty. Also note that $\mathcal{M}_\infty \subset \mathcal{M}_r$ for all $r > 0$. The most prominent aspects in quantization of probability measures are the following:

Find the exact value of the upper/lower quantization dimension for μ of order r: In the case where the quantization dimension does not exist, it is usually difficult to obtain the exact value of the upper or lower one (cf. [30]). Up to now, in such a situation, the upper and lower quantization dimension could only be explicitly determined for very special cases.

Determine the s-dimensional upper and lower quantization coefficient: We are mainly concerned about the finiteness and positivity of these quantities. This question is analogous to the question of whether a fractal is an s-set. Typically, this question is much harder to answer than finding the quantization dimension. So far, the quantization coefficient has been studied for absolutely continuous probability measures ([6]) and several classes of singular measures, including self-similar and self-conformal [19, 29, 33, 39] measures, Markov-type measures [16, 29, 44] and self-affine measures on Bedford-McMullen carpets [15, 38].

Properties of the point density measure $\boldsymbol{\mu}_{\mathbf{r}}$: Fix a sequence of n-optimal sets $\alpha_n \in C_{n,r}(\mu)$ of order r, $n \in \mathbb{N}$, and consider the weak limit of the empirical measures, whenever it exists,

$$\mu_r := \lim_{n\to\infty} \frac{1}{n} \sum_{a\in\alpha_n} \delta_a.$$

The point density measure characterizes the frequency at which optimal points fall into a given open set. Up to now, the point density measure is determined only for absolutely continuous measures [6, Theorem 7.5] and certain self-similar measures [9, Theorem 5.5].

Local properties and Voronoi partitions: Fix a finite subset α of $\mathbb{R}^q$. A Voronoi partition with respect to α refers to a partition $(P_a(\alpha))_{a\in\alpha}$ of $\mathbb{R}^q$ such that

$$P_a(\alpha) \subset \{x \in \mathbb{R}^q : d(x,\alpha) = d(x,a)\}, \ a \in \alpha.$$

It is natural to ask, if there exists constants $0 < C_1 \leq C_2 < \infty$ such that for all $\alpha_n \in C_{n,r}(\mu)$ and $n \in \mathbb{N}$ we have

$$\frac{C_1 e^r_{n,r}}{n} \leq \min_{a\in\alpha_n} \int_{P_a(\alpha_n)} \mathrm{d}(x,\alpha_n)^r \,\mathrm{d}\mu(x)$$
$$\leq \max_{a\in\alpha_n} \int_{P_a(\alpha_n)} \mathrm{d}(x,\alpha_n)^r \,\mathrm{d}\mu(x) \leq \frac{C_2 e^r_{n,r}}{n}.$$

This question is essentially a weaker version of Gersho's conjecture [5]. Graf, Luschgy and Pagès proved in [10] that this is in fact true for a large class of absolutely continuous probability measures. An affirmative answer is also given for self-similar measures under the assumption of the strong separation condition (SSC) for the corresponding iterated function system [39, 43].

In the final analysis, the study of the quantization problem addresses the optimal sets. Where do the points of an optimal set lie? Unfortunately, it is almost impossible to determine the optimal sets for a general probability measure. It is therefore desirable to seek for an "approximately explicit" description of such sets. In other words, even though we do not know exactly where the points of an optimal set lie, we want to know how many points are lying in a given open set. This would in return enable us to obtain precise estimates for the quantization error.

Connection to fractal geometry: To this end, some typical techniques in fractal geometry are often very helpful. In fact, the quantization problem is closely connected with important notions in fractal geometry. One may compare the upper (lower) quantization dimension for measures to the packing (Hausdorff) dimension for sets; accordingly, the upper (lower) quantization coefficient may be compared to the packing (Hausdorff) measure for sets. Although they are substantially different, they do have some close connections, as all these quantities can be defined in terms of coverings, partitions and packings. In fact, we have

(1) $\dim^*_H \mu \leq \underline{D}_r(\mu) \leq \underline{\dim}^*_B\mu$ and $\dim^*_P \mu \leq \overline{D_r}(\mu) \leq \overline{\dim}^*_B\mu$, for $r = 2$ these inequalities were presented in [26], and for measures with compact support and all $r \in (0, \infty]$ they were independently proved in [6].

(2) In [14] we have studied the *stability* of the upper and lower quantization dimension in some detail. In [14], for $r \in [1, \infty]$, we proved the following statements:

 (i) For all $\mu \in \mathcal{M}_r$ we have $\overline{D}_r(\mu) = \max_{1\leq i\leq n} \overline{D}_r(\mu_i)$ with $\mu_i \in \mathcal{M}_r$, $s_i > 0$, $i = 1, \ldots, n$, $n \in \mathbb{N}$ and $\mu = \sum_{i=1}^n s_i\mu_i$.

 (ii) $\dim^*_P(\mu) = \inf\left\{\sup_{i\in\mathbb{N}} \overline{D}_r(\mu_i) : \mu = \sum_{i\in\mathbb{N}} s_i\mu_i, \mu_i \in \mathcal{M}_\infty, s_i > 0\right\}$ for all $\mu \in \mathcal{M}_\infty$.

 (iii) There exists $\mu \in \mathcal{M}_\infty$ such that $\overline{D}_r(\mu) \neq \dim^*_P(\mu)$.

(iv) There exists $\mu \in \mathcal{M}_\infty$ such that $\overline{D}_r(\mu) > \underline{D}_r(\mu)$.
(v) There exists $\mu \in \mathcal{M}_\infty$ such that $\underline{D}_r(\mu) \neq \max_{1\leq i\leq n} \underline{D}_r(\mu_i)$ for some $\mu_i \in \mathcal{M}_\infty, s_i > 0, i = 1, \ldots, n, n \in \mathbb{N}$ with $\mu = \sum_{i=1}^{n} s_i\mu_i$.

(3) For certain measures arising from dynamical systems, the quantization dimension can be expressed within the thermodynamic formalism in terms of appropriate *temperature functions* (see [15, 19, 27, 28]).
(4) The upper and lower quantization dimension of order zero are closely connected with the *upper and lower local dimension*. As it is shown in [43], if ν-almost everywhere the upper and lower local dimension are both equal to s, then $D_0(\nu)$ exists and equals s.

We end this section with Graf and Luschgy's results on self-similar measures. These results and the methods involved in their proofs have a significant influence on subsequent work on the quantization for non-self-similar measures.

Let $(S_i)_{i=1}^N$ be a family of contractive similitudes on $\mathbb{R}^q$ with contraction ratios $(s_i)_{i=1}^N$. According to [12], there exists a unique non-empty compact subset E of $\mathbb{R}^q$ such that $E = \bigcup_{i=1}^N S_i(E)$. The set E is called the self-similar set associated with $(S_i)_{i=1}^N$. Also, there exists a unique Borel probability measure on $\mathbb{R}^q$, such that $\mu = \sum_{i=1}^N p_i\mu \circ S_i^{-1}$, called the self-similar measure associated with $(S_i)_{i=1}^N$ and the probability vector $(p_i)_{i=1}^N$. We say that $(S_i)_{i=1}^N$ satisfies the *strong separation condition* (SSC) if the sets $S_i(E), i = 1, \cdots, N$, are pairwise disjoint. We say that it satisfies the *open set condition* (OSC) if there exists a non-empty open set U such that $S_i(U) \cap S_j(U) = \emptyset$ for all $i \neq j$ and $S_i(U) \subset U$ for all $i = 1, \cdots, N$. For $r \in [0, \infty)$, let k_r be the positive real number given by

$$k_0 := \frac{\sum_{i=1}^N p_i \log p_i}{\sum_{i=1}^N p_i \log s_i}, \quad \sum_{i=1}^N (p_i c_i^r)^{\frac{k_r}{k_r+r}} = 1. \tag{1.2}$$

Theorem 1.2 ([7, 8]) *Assume that $(S_i)_{i=1}^N$ satisfies the open set condition. Then for all $r \in [0, \infty)$, we have*

$$0 < \underline{Q}_r^{k_r}(\mu) \leq \overline{Q}_r^{k_r}(\mu) < \infty.$$

In particular, we have $D_r(\mu) = k_r$.

This is the first complete result on the quantization for (typically) singular measures. In its proof, Hölder's inequality with an exponent less than one plays a crucial role, from which the exponent $k_r/\,(k_r + r)$ comes out in a natural manner.

2 The Three-Step Procedure

Following the ideas of Graf-Luschgy we propose a three-step procedure for the estimation of the quantization errors by means of partitions, coverings and packings. This procedure is applicable to a large class of fractal measures, including Moran measures, self-affine measures and Markov-type measures, provided that some suitable separation condition is satisfied; it even allows us to obtain useful information on the quantization for general Borel probability measures on $\mathbb{R}^q$ with compact support.

Step 1 (Partitioning). For each n, we partition the (compact) support of μ into φ_n small parts $(F_{nk})_{k=1}^{\varphi_n}$, such that $\mu(F_{nk})|F_{nk}|^r$ are uniformly comparable, namely, for some constant $C > 1$ independent of $k, j \in \{1, \dots, \varphi_n\}$ and $n \in \mathbb{N}$, we have

$$C^{-1}\mu(F_{nk})|F_{nk}|^r \leq \mu(F_{nj})|F_{nj}|^r \leq C\mu(F_{nk})|F_{nk}|^r,$$

where $|A|$ denotes the diameter of a set $A \subset \mathbb{R}^d$. This idea was first used by Graf and Luschgy to treat the quantization problem for self-similar measures, we refer to [6] for a construction of this type. The underlying idea is to seek for some uniformity while μ generally is not uniform.

Step 2 (Covering). With a suitable separation condition, we may also assume that for some $\delta > 0$, we have that

$$d(F_{nk}, F_{nj}) \geq \delta \max\{|F_{nk}|, |F_{nj}|\},\ k \neq j,\ n \geq 1.$$

In this step, uniformity and separation allow us to verify that any φ_n-optimal set distributes its points equally among suitable neighborhoods of F_{nk}, $1 \leq k \leq \varphi_n$, in other words, each F_{nk} "owns" a bounded number of points of the φ_n-optimal set. More precisely, we prove that there exists some constant L_1, which is independent of n, such that for every $\alpha \in C_{\varphi_n,r}(\mu)$, we have

$$\max_{1\leq k\leq\varphi_n} \operatorname{card}\left(\alpha \cap (F_{nk})_{4^{-1}\delta|F_{nk}|}\right) \leq L_1,$$

where A_s denotes the *s-parallel set* of A. This can often be done inductively by means of contradiction.

Step 3 (Packing). In the last step we have to find a constant L_2 and subsets β_{nk} of F_{nk} with cardinality at most L_2 such that for all $\alpha \in C_{\varphi_n,r}(\mu)$ and $x \in F_{nk}$ we have

$$d(x, \alpha) \geq d(x, (\alpha \cap (F_{nk})_{4^{-1}\delta|F_{nk}|}) \cup \beta_{nk}).$$

This reduces the global situation to a local one and enables us to restrict our attention to an arbitrary small set F_{nk}. We have

$$e^r_{\varphi_n,r}(\mu) \geq \sum_{k=1}^{\varphi_n} \int_{F_{nk}} d(x, (\alpha \cap (F_{nk})_{4^{-1}\delta|F_{nk}|}) \cup \beta_{nk})^r \, \mathrm{d}\mu(x).$$

Note that card $\big((\alpha \cap (F_{nk})_{4^{-1}\delta|F_{nk}|}) \cup \beta_{nk}\big) \leq L_1 + L_2$. For measures with explicit mass distributions, we often have

$$\int_{F_{nk}} d(x, \gamma \cup \beta_k)^r \, \mathrm{d}\mu(x) \geq D\mu(F_{nj})|F_{nj}|^r$$

for any subset γ of $\mathbb{R}^q$ with cardinality not greater than $L_1 + L_2$ and an appropriate constant D. Thus, we get a lower estimate for the quantization error:

$$e^r_{\varphi_n,r}(\mu) \geq D \sum_{k=1}^{\varphi_n} \mu(F_{nk})|F_{nk}|^r.$$

On the other hand, by choosing some arbitrary points $b_k \in F_{nk}$, $k \in \{1, \ldots, \varphi_n\}$, one can easily see

$$e^r_{\varphi_n,r}(\mu) \leq \sum_{k=1}^{\varphi_n} \int_{F_{nk}} d(x, b_k)^r d\mu(x) \leq \sum_{k=1}^{\varphi_n} \mu(F_{nk})|F_{nk}|^r.$$

After these three steps, for sufficiently "nice" measures, we may additionally assume that $\varphi_n \leq \varphi_{n+1} \leq C\varphi_n$ for some constant $C > 1$ (cf. [34–36]). To determine the dimension it is then enough to estimate the growth rate of φ_n. Here, ideas from Thermodynamic Formalism – such as critical exponents or zeros of some pressure function – often come into play: E.g., for $r > 0$ we often have

$$\frac{D_r(\mu)}{D_r(\mu) + r} = \inf \left\{ t \in \mathbb{R} : \sum_{n \in \mathbb{N}} \sum_{k=1}^{\varphi_n} (\mu(F_{nk})|F_{nk}|^r)^t < \infty \right\}$$

allowing us to find explicit formulae for the quantization dimension for a given problem (see [15] for an instance of this). Typically, for a non-self-similar measure such as a self-affine measures on Bedford-McMullen carpets, this requires a detailed analysis of the asymptotic quantization errors. In order to formulate a rigorous proof, we usually need to make more effort according to the particular properties of the measures under consideration. As general measures do not enjoy strict self-similarity, it seems unrealistic to expect to establish simple quantities for the quantization errors as Graf and Luschgy did for self-similar measures [6, Lemma 14.10]. However, the above-mentioned three-step procedure often provides us with estimates of the quantization errors which is usually a promising starting point.

Moreover, in order to examine the finiteness or positivity of the upper and lower s-dimensional quantization coefficient of order r, it suffices to check that (cf. [40])

$$0 < \liminf_{n\to\infty} \sum_{k=1}^{\varphi_n} (\mu(F_{nj})|F_{nj}|^r)^{\frac{s}{s+r}} \leq \limsup_{n\to\infty} \sum_{k=1}^{\varphi_n} (\mu(F_{nj})|F_{nj}|^r)^{\frac{s}{s+r}} < \infty.$$

An effective way to do this is to construct some auxiliary probability measures. Such a measure should closely reflect the information carried by $(\mu(F_{nj})|F_{nj}|^r)^{\frac{s}{s+r}}$. For a self-similar measure, as Graf-Luschgy's work shows, an auxiliary probability measure is the self-similar measure associated with $(S_i)_{i=1}^N$ and the probability vector $((p_i c_i^r)^{\frac{k_r}{k_r+r}})_{i=1}^N$. It is interesting to note that this measure coincides with the point density measure provided that the k_r-dimensional quantization coefficient exists. For a self-similar measure, as Graf and Luschgy showed, we can use the above auxiliary probability measure and obtain the finiteness or positivity of the upper and lower k_r-dimensional quantization coefficient, which also implies that the quantization dimension exists and equals k_r. In the non-self-similar situation, due to the complexity of the topological support, it is often not easy to construct a suitable auxiliary probability measure to estimate the quantization coefficients.

3 Recent Work on the Quantization for Fractal Measures

3.1 Self-Affine Measures on Bedford-McMullen Carpets

Fix two positive integers m, n with $2 \leq m \leq n$ and fix a set

$$G \subset \{0, 1, \ldots, n-1\} \times \{0, 1, \ldots, m-1\}$$

with $N := \text{card}\,(G) \geq 2$. We define a family of affine mappings on $\mathbb{R}^2$ by

$$f_{ij} : (x, y) \mapsto \left(n^{-1}x + n^{-1}i, m^{-1}y + m^{-1}j\right), \quad (i, j) \in G. \tag{3.1}$$

By [12], there exists a unique non-empty compact set E satisfying $E = \bigcup_{(i,j)\in G} f_{ij}(E)$, which is called the Bedford-McMullen carpet determined by $(f_{ij})_{(i,j)\in G}$. Given a probability vector $(p_{ij})_{(i,j)\in G}$ with $p_{ij} > 0$, for all $(i, j) \in G$, the self-affine measure associated with $(p_{ij})_{(i,j)\in G}$ and $(f_{ij})_{(i,j)\in G}$ refers to the unique Borel probability measure μ on $\mathbb{R}^2$ satisfying

$$\mu = \sum_{(i,j)\in G} p_{ij}\mu \circ f_{ij}^{-1}. \tag{3.2}$$

Sets and measures of this form have been intensively studied in the past decades, see e.g. [1, 4, 11, 17, 18, 21, 25] for many interesting results. We write

$$G_x := \{i : (i,j) \in G \text{ for some } j\}, \; G_y := \{j : (i,j) \in G \text{ for some } i\},$$
$$G_{x,j} := \{i : (i,j) \in G\}, \;\; q_j := \sum_{i \in G_{x,j}} p_{ij}.$$

We carry out the three-step procedure and obtain an estimate for the quantization errors. This allows us to conjecture that the quantization dimension exists and equals s_r, where

$$\left(\sum_{(i,j)\in G} (p_{ij} m^{-r})^{\frac{s_r}{s_r+r}} \right)^{\theta} \left(\sum_{j \in G_y} (q_j m^{-r})^{\frac{s_r}{s_r+r}} \right)^{1-\theta} = 1, \; \theta := \frac{\log m}{\log n}. \tag{3.3}$$

However, it seems rather difficult to find a suitable auxiliary measure for a proof of this conjecture. A cornerstone is the crucial observation that the number s_r coincide with a Poincare-like exponent [15]. Using the property of sup-additive sequences, we are able to prove that $D_r(\mu)$ exists and also coincides with κ_r. Finally, we consider the self-affine measure associated with $((p_{ij} m^{-r})^{\frac{s_r}{s_r+r}}/C_r)_{(i,j)\in G}$ as an auxiliary measure, where $C_r := \sum_{(i,j)\in G} (p_{ij} m^{-r})^{\frac{s_r}{s_r+r}}$. This measure and the above-mentioned estimate enable us to obtain sufficient conditions for the upper and lower quantization coefficient to be both positive and finite. We have

Theorem 3.1 ([15]) *Let μ be as defined in (3.2). Then for each $r \in (0, \infty)$ we have that $D_r(\mu)$ exists and equals s_r, Moreover, $0 < \underline{Q}_r^{s_r}(\mu) \le \overline{Q}_r^{s_r}(\mu) < \infty$ if one of the following conditions is fulfilled:*

(A) $\sum_{i \in G_{x,j}} (p_{ij} q_j^{-1})^{\frac{s_r}{s_r+r}}$ *are identical for all $j \in G_y$,*
(B) q_j *are identical for all $j \in G_y$.*

Open problem: Is it true that $0 < \underline{Q}_r^{s_r}(\mu) \le \overline{Q}_r^{s_r}(\mu) < \infty$ if and only if condition (A) or (B) holds?

3.2 Quantization for Markov-Type Measures

3.2.1 Mauldin-Williams Fractals

Let J_i, non-empty compact subsets of $\mathbb{R}^d$ with $J_i = \mathrm{cl}(\mathrm{int}(J_i))$, $1 \le i \le N$, where $\mathrm{cl}(A)$ and $\mathrm{int}(A)$ denote the closure and interior in $\mathbb{R}^d$ of a set $A \subset \mathbb{R}^d$. For the integer $N \ge 2$ let $P = (p_{ij})_{1 \le i,j \le N}$ be a row-stochastic matrix, i.e., $p_{ij} \ge 0$, $1 \le i,j \le N$, and $\sum_{j=1}^{N} p_{ij} = 1, 1 \le i \le N$. Let θ denote the empty word and set

$$\Omega_0 := \{\theta\},\ \Omega_1 := \{1, \dots N\},$$
$$\Omega_k := \{\sigma \in \Omega_1^k : p_{\sigma_1\sigma_2} \cdots p_{\sigma_{k-1}\sigma_k} > 0\},\ k \ge 2,$$
$$\Omega^* := \bigcup_{k\ge 0} \Omega_k,\ \Omega_\infty := \{\sigma \in \Omega_1^{\mathbb{N}} : p_{\sigma_h\sigma_{h+1}} > 0 \text{ for all } h \ge 1\}.$$

We call J_i, $1 \le i \le N$, cylinder sets of order one. For each $1 \le i \le N$, let $J_{ij},\ (i,j) \in \Omega_2$, be non-overlapping subsets of J_i such that J_{ij} is geometrically similar to J_j and $\text{diam}(J_{ij})/\text{diam}(J_j) = c_{ij}$. We call these sets cylinder sets of order two. Assume that cylinder sets of order k are determined, namely, for each $\sigma \in \Omega_k$, we have a cylinder set J_σ. Let $J_{\sigma * i_{k+1}}, \sigma * i_{k+1} \in \Omega_{k+1}$, be non-overlapping subsets of J_σ such that $J_{\sigma * i_{k+1}}$ is geometrically similar to $J_{i_{k+1}}$. Inductively, cylinder sets of order k are determined for all $k \ge 1$. The *(ratio specified) Mauldin-Williams fractal* is given by

$$E := \bigcap_{k\ge 1} \bigcup_{\sigma\in\Omega_k} J_\sigma.$$

3.2.2 Markov-Type Measures

Let $(\chi_i)_{i=1}^N$ be an arbitrary probability vector with $\min_{1\le i\le N} \chi_i > 0$. By Kolmogorov consistency theorem, there exists a unique probability measure $\tilde{\mu}$ on Ω_∞ such that $\tilde{\mu}([\sigma]) := \chi_{\sigma_1} p_{\sigma_1\sigma_2} \cdots p_{\sigma_{k-1}\sigma_k}$ for every $k \ge 1$ and $\sigma = (\sigma_1, \dots, \sigma_k) \in \Omega_k$, where $[\sigma] := \{\omega \in \Omega_\infty : \omega|_{|\sigma|} = \sigma\}$. Let π denote the projection from Ω_∞ to E given by $\pi(\sigma) := x$, where

$$\{x\} := \bigcap_{k\ge 1} J_{\sigma|_k}, \quad \text{for } \sigma \in \Omega_\infty.$$

Let us assume the following:

(A1) card($\{j : p_{ij} > 0\}) \ge 2$ for all $1 \le i \le N$.
(A2) There exists a constant $t \in (0, 1)$ such that for every $\sigma \in \Omega^*$ and distinct $i_1, i_2 \in \Omega_1$ with $\sigma * i_l \in \Omega_{|\sigma|+1}$, $l = 1, 2$,

$$d(J_{\sigma * i_1}, J_{\sigma * i_2}) \ge t \max\{|J_{\sigma * i_1}|, |J_{\sigma * i_2}|\}.$$

Under this assumption, π is a bijection. We consider the image measure of $\tilde{\mu}$ under the projection π given by $\mu := \tilde{\mu} \circ \pi^{-1}$. We call μ a Markov-type measure which satisfies

$$\mu(J_\sigma) = \chi_{\sigma_1} p_{\sigma_1\sigma_2} \cdots p_{\sigma_{k-1}\sigma_k} \text{ for } \sigma = (\sigma_1 \dots \sigma_k) \in \Omega_k. \tag{3.4}$$

For $1 \leq i,j \leq N$, we define $a_{ij}(s) := (p_{ij}c_{ij}^r)^s$. Then we get an $N \times N$ matrix $A(s) = (a_{ij}(s))_{N\times N}$. Let $\psi(s)$ denote the spectral radius of $A(s)$. By [20, Theorem 2], $\psi(s)$ is continuous and strictly decreasing. Note that, by the assumption (A1), the Perron-Frobenius theorem and intermediate-value theorem, there exists a unique number $\xi \in (0,1)$ such that $\psi(\xi) = 1$. Thus, for every $r > 0$, there exists a unique positive number s_r such that $\psi(\frac{s_r}{s_r+r}) = 1$.

We consider the directed graph G associated with the transition matrix $(p_{ij})_{N\times N}$. Namely, G has vertices $1, 2, \ldots, N$. There is an edge from i to j if and only if $p_{ij} > 0$. In the following, we will simply denote by $G = \{1, \ldots, N\}$ both the directed graph and its vertex sets. We also write

$$b_{ij}(s) := (p_{ij}c_{ij}^r)^{\frac{s}{s+r}}, \quad A_{G,s} := (b_{ij}(s))_{N\times N}, \quad \Psi_G(s) := \psi\left(\frac{s}{s+r}\right).$$

Let $\mathrm{SC}(G)$ denote the set of all strongly connected components of G. For $H_1, H_2 \in \mathrm{SC}(G)$, we write $H_1 \prec H_2$, if there is a path initiating at some $i_1 \in H_1$ and terminating at some $i_k \in H_2$. If we have neither $H_1 \prec H_2$ nor $H_2 \prec H_1$, then we say H_1, H_2 are incomparable. For every $H \in \mathrm{SC}(G)$, we denote by $A_{H,s}$ the sub-matrix $(b_{ij}(s))_{i,j\in H}$ of $A_G(s)$. Let $\Psi_H(s)$ be the spectral radius of $A_{H,s}$ and $s_r(H)$ be the unique positive number satisfying $\Psi_H(s_r(H)) = 1$.

Again, we apply the three-step procedure in Sect. 2 and obtain upper and lower estimates for the quantization error. Using these estimates and auxiliary measures of Mauldin-Williams type, we are able to prove that, when the transition matrix is irreducible, the upper and lower quantization coefficient are both positive and finite. This fact also leads to the positivity of the lower quantization coefficient in the general case. Then, based on a detailed analysis of the corresponding directed graph (not strongly connected) and some techniques in matrix theory, we are able to prove the formula for the quantization dimension. Finally, by using auxiliary measures of Mauldin-Williams type once more, we establish a necessary and sufficient condition for the upper quantization coefficient to be finite as stated next.

Theorem 3.2 ([16]) *Assume that (A1) and (A2) are satisfied. Let μ be the Markov-type measure as defined in (3.4) and s_r the unique positive number satisfying $\Psi_G(s_r) = 1$. Then, $D_r(\mu) = s_r$ and $\underline{Q}_r^{s_r}(\mu) > 0$. Furthermore, $\overline{Q}_r^{s_r}(\mu) < \infty$ if and only if $\mathcal{M} := \{H \in \mathrm{SC}(G) : s_r(H) = s_r\}$ consists of incomparable elements, otherwise, we have $\underline{Q}_r^{s_r}(\mu) = \infty$.*

3.3 *Quantization for Moran Measures*

3.3.1 Moran Sets

Let J be a non-empty compact subset of $\mathbb{R}^d$ with $J = \mathrm{cl}(\mathrm{int}(J))$. Let $|A|$ denote the diameter of a set $A \subset \mathbb{R}^d$. Let $(n_k)_{k=1}^{\infty}$ be a sequence of integers with $\min_{k\geq 1} n_k \geq 2$

and θ denote the empty word. Set

$$\Omega_0 := \{\theta\},\ \Omega_k := \prod_{j=1}^{k}\{1,2,\cdots,n_j\},\ \Omega^* := \bigcup_{k=0}^{\infty}\Omega_k.$$

For $\sigma = \sigma_1\cdots\sigma_k \in \Omega_k$ and $j \in \{1,\cdots,n_{k+1}\}$, we write $\sigma * j = \sigma_1\cdots\sigma_k j$.

Set $J_\theta := J$ and let J_σ for $\sigma \in \Omega_1$ be non-overlapping subsets of J_θ such that each of them is geometrically similar to J_θ. Assume that J_σ is determined for every $\sigma \in \Omega_k$. Let $J_{\sigma*j}, 1 \le j \le n_{k+1}$ be non-overlapping subsets of J_σ which are geometrically similar to J_σ. Inductively, all sets $J_\sigma, \sigma \in \Omega^*$ are determined in this way. The Moran set is then defined by

$$E := \bigcap_{k=1}^{\infty}\bigcup_{\sigma\in\Omega_k} J_\sigma. \tag{3.5}$$

We call $J_\sigma, \sigma \in \Omega_k$, cylinders of order k. It is well known that the Moran sets E are generally not self-similar (cf. [3, 31]). For $k \ge 0$ and $\sigma \in \Omega_k$, we set

$$|\sigma| := k,\ c_{\sigma,j} := \frac{|J_{\sigma*j}|}{|J_\sigma|},\ 1 \le j \le n_{k+1}.$$

We assume that there exist some constants $c, \beta \in (0,1)$ such that

(B1) $\inf_{\sigma\in\Omega^*}\min_{1\le j\le n_{|\sigma|+1}} c_{\sigma,j} = c > 0,$

(B2) $\mathrm{dist}(J_{\sigma*i}, J_{\sigma*j}) \ge \beta \max\{|J_{\sigma*i}|, |J_{\sigma*j}|\}$ for $1 \le i \ne j \le n_{|\sigma|+1}$ and $\sigma \in \Omega^*$.

3.3.2 Moran Measures

For each $k \ge 1$, let $(p_{kj})_{j=1}^{n_k}$ be a probability vector. By the Kolmogorov consistency theorem, there exists a probability measure ν on $\Omega_\infty := \prod_{k=1}^{\infty}\{1,2,\cdots,n_k\}$ such that

$$\nu([\sigma_1,\cdots,\sigma_k]) = p_{1\sigma_1}\cdots p_{k\sigma_k},\ \sigma_1\cdots\sigma_k \in \Omega_k,$$

where $[\sigma_1,\cdots,\sigma_k] = \{\tau \in \Omega_\infty : \tau_j = \sigma_j, 1 \le j \le k\}$. Let $\Pi : \Omega_\infty \to E$ be defined by $\Pi(\sigma) = \bigcap_{k\ge1} J_{\sigma|_k}$ with $\sigma|_k = \sigma_1\cdots\sigma_k$. Then, with the assumption (B2), Π is a continuous bijection. We define $\mu := \nu\circ\Pi^{-1}$. Then, we have

$$\mu(J) = 1,\ \mu(J_\sigma) := p_{1\sigma_1}\cdots p_{k\sigma_k},\ \ \sigma = \sigma_1\cdots\sigma_k \in \Omega_k,\ k \ge 1.$$

We call the measure μ the Moran measure on E. It is known that the quantization dimension for μ of order r does not necessarily exist. Let $d_{k,r}, \overline{d}_r, \underline{d}_r$ be given by

$$\sum_{\sigma\in\Omega_k}(p_\sigma c_\sigma^r)^{\frac{d_{k,r}}{d_{k,r}+r}} = 1,\ \ \overline{d}_r := \limsup_{k\to\infty} d_{k,r},\ \ d_{k,r}\underline{d}_r := \liminf_{k\to\infty} d_{k,r}.$$

Open problem Is it true that $\overline{D}_r(\mu) = \overline{d}_r, \underline{D}_r(\mu) = \underline{d}_r$?

3.3.3 Multiscale Moran Sets

A multiscale Moran set is Moran set with some additional structure encoded in the infinite sequence $\omega = (\omega_l)_{l=1}^{\infty} \in \Upsilon := \{1,\ldots,m\}^{\mathbb{N}}$ for some $m \geq 2$. For this fix some positive integers $N_i \geq 2, 1 \leq i \leq m$ and for every $1 \leq i \leq m$, let $(g_{ij})_{j=1}^{N_i}$ be the contraction vector with $g_{ij} \in (0,1)$ and $(p_{ij})_{j=1}^{N_i}$ a probability vector with $p_{ij} > 0$ for all $1 \leq j \leq N_i$. Now using the notation in the definition of Moran sets, we set

$$n_{l+1} := N_{\omega_{l+1}},\ \ \left(c_{\sigma,j}\right)_{j=1}^{N_{\omega_{l+1}}} := (g_{\omega_{l+1}j})_{j=1}^{N_{\omega_{l+1}}},\ \sigma \in \Omega_l,\ l \geq 0. \tag{3.6}$$

If, for some $l \geq 0$, we have $\omega_{l+1} = i$, then for every $\sigma \in \Omega_l$, we have a continuum of choices of $\{J_{\sigma * j}\}_{j=1}^{N_i}$ fulfilling (B1),(B2) and (3.6), because we only fix the contraction ratios of the similitudes. Hence, to every $\omega \in \Upsilon$, there corresponds a class $\mathcal{M}_\omega$ of Moran sets according to (3.5). We call these Moran sets *multiscale Moran sets*.

For each $\omega \in \Upsilon$, we write

$$N_{k,i}(\omega) := \text{card}\{1 \leq l \leq k : \omega_l = i\},\ \ 1 \leq i \leq m.$$

Fix a probability vector $\chi = (\chi_i)_{i=1}^m$ with $\chi_i > 0$ for all $1 \leq i \leq m$ and define

$$G(\chi) := \{\omega \in \Upsilon : \lim_{k\to\infty} k^{-1} N_{k,i}(\omega) = \chi_i,\ 1 \leq i \leq m\},$$

$$G_0(\chi) := \{\omega \in \Upsilon : \limsup_{k\to\infty} \left|N_{k,i}(\omega) - k\chi_i\right| < \infty, 1 \leq i \leq m\}.$$

3.3.4 Multiscale Moran Measures

Fix an $\omega \in G(\chi)$. According to Kolmogorov consistency theorem, there exists a probability measure ν_ω on the product space $\Omega_\infty := \prod_{k=1}^{\infty}\{1,2,\cdots,N_{\omega_k}\}$ such that

$$\nu_\omega([\sigma_1,\cdots,\sigma_k]) = p_{\omega_1\sigma_1}\cdots p_{\omega_k\sigma_k},\ \sigma_1\cdots\sigma_k \in \Omega_k,$$

where $[\sigma_1, \cdots, \sigma_k] = \{\tau \in \Omega_\infty : \tau_j = \sigma_j, 1 \le j \le k\}$. We define $\mu_\omega := \nu_\omega \circ \Pi^{-1}$. Then, we have

$$\mu(J) = 1,\ \mu_\omega(J_\sigma) := p_{\omega_1\sigma_1} \cdots p_{\omega_k\sigma_k},\quad \sigma = \sigma_1 \cdots \sigma_k \in \Omega_k,\ k \ge 1.$$

We call the measure μ_ω the infinite product measure on $E(\omega)$ associated with ω and $(p_{ij})_{j=1}^{N_i}, 1 \le i \le m$.

For every $\omega \in G(\chi)$ and $k \in \mathbb{N}$, let $s_{k,r}(\omega), s_r$ and $\underline{H}_r(\omega), \overline{H}_r(\omega)$ be defined by

$$\prod_{i=1}^{m}\left(\sum_{j=1}^{N_i}(p_{ij}g_{ij}^r)^{\frac{s_{k,r}(\omega)}{s_{k,r}(\omega)+r}}\right)^{N_{k,i}(\omega)} = 1,\ \ \prod_{i=1}^{m}\left(\sum_{j=1}^{N_i}(p_{ij}g_{ij}^r)^{\frac{s_r}{s_r+r}}\right)^{\chi_i} = 1, \tag{3.7}$$

$$\underline{H}_r(\omega) := \liminf_{k\to\infty} k|s_{k,r}(\omega) - s_r|,\ \ \overline{H}_r(\omega) := \limsup_{k\to\infty} k|s_{k,r}(\omega) - s_r|.$$

Compared with Mauldin-Williams fractals, the disadvantage is that we have more patterns in the construction of multiscale Moran sets. However, the pattern we use at the $(k+1)$-th step is independent of words of length k, which is an advantage. After we carry out the three-step procedure in Sect. 2, we conveniently obtain the exact value of the quantization dimension by considering some measure-like auxiliary functions. This also enables us to transfer the question of the upper and lower quantization coefficient to the convergence order of $(s_{k,r}(\omega))_{k=1}^{\infty}$. For the latter, we need a detailed analysis of some auxiliary functions related to (3.7). One may see [40] for more details. Our main result is summarized in the following theorem.

Theorem 3.3 ([40]) *For every $\omega \in G(\chi)$, we have*

(i) *$D_r(\mu_\omega)$ exists and equals s_r, it is independent of $\omega \in G(\chi)$,*
(ii) *If $s_{k,r}(\omega) \ge s_r$ for all large k, then $\underline{Q}_r^{s_r}(\mu_\omega) > 0$. If in addition $\underline{H}_r(\omega) = \infty$, then we have $\overline{Q}_r^{s_r}(\mu_\omega) = \infty$,*
(iii) *If $s_{k,r}(\omega) \le s_r$ for all large k, then $\overline{Q}_r^{s_r}(\mu_\omega) < \infty$; if, in addition, $\underline{H}_r(\omega) = \infty$, then we have $\underline{Q}_r^{s_r}(\mu_\omega) = 0$,*
(iv) *If $\overline{H}_r(\omega) < \infty$, then $\underline{Q}_r^{s_r}(\mu_\omega)$ and $\overline{Q}_r^{s_r}(\mu_\omega)$ are both positive and finite,*
(v) *If $\omega \in G_0(\chi)$, then the assertion in* (iv) *holds.*

Open problem: What can we say about necessary conditions for $\underline{Q}_r^{s_r}(\mu_\omega)$ and $\overline{Q}_r^{s_r}(\mu_\omega)$ to be both positive and finite?

References

1. T. Bedford, Crinkly curves, Markov partitions and box dimensions in self-similar sets, Ph.D. thesis, University of Warwick (1984)
2. J.A. Bucklew, G.L. Wise, Multidimensional asymptotic quantization with rth power distortion measures. IEEE Trans. Inform. Theory **28**, 239–247 (1982)

3. R. Cawley, R.D. Mauldin, Multifractal decompositions of Moran fractals. Adv. Math. **92**, 196–236 (1992)
4. K.J. Falconer, Generalized dimensions of measures on almost self-affine sets. Nonlinearity **23**, 1047–1069 (2010)
5. A. Gersho, Asymptotically optimal block quantization. IEEE Trans. Inform. Theory **25**, 373–380 (1979)
6. S. Graf, H. Luschgy, *Foundations of Quantization for Probability Distributions*. Lecture Notes in Mathematics, vol. 1730. (Springer, Berlin/New York, 2000)
7. S. Graf, H. Luschgy, Asymptotics of the quantization error for self-similar probabilities. Real. Anal. Exch. **26**, 795–810 (2001)
8. S. Graf, H. Luschgy, Quantization for probabilitiy measures with respect to the geometric mean error. Math. Proc. Camb. Phil. Soc. **136**, 687–717 (2004)
9. S. Graf, H. Luschgy, The point density measure in the quantization of self-similar probabilities. Math. Proc. Camb. Phil. Soc. **138**, 513–531 (2005)
10. S. Graf, H. Luschgy, G. Pagès, The local quantization behavior of absolutely continuous probabilities. Ann. Probab. **40**(4), 1795–1828 (2012)
11. Y. Gui, W.X. Li, Multiscale self-affine Sierpinski carpets. Nonlinearity **23**, 495–512 (2010)
12. J.E. Hutchinson, Fractals and self-similarity. Indiana Univ. Math. J. **30**, 713–747; 7–156 (1981)
13. M. Kesseböhmer, S. Zhu, Quantization dimension via quantization numbers. Real Anal. Exch. **29**(2), 857–866 (2003/2004)
14. M. Kesseböhmer, S. Zhu, Stability of quantization dimension and quantization for homogeneous Cantor measures. Math. Nachr. **280**(8), 866–881 (2007)
15. M. Kesseböhmer, S. Zhu, On the quantization for self-affine measures on Bedford-McMullen carpets. arXiv:1312.3289 (2013)
16. M. Kesseböhmer, S. Zhu, The quantization for Markov-type measures on a class of ratio-specified graph directed fractals. preprint arXiv:1406.3257 (2014)
17. J.F. King, The singularity spectrum for general Sierpiński carpets. Adv. Math. **116**, 1–11 (1995)
18. S.P. Lalley, D. Gatzouras, Hausdorff and box dimensions of certain self-affine fractals. Indiana Univ. Math. J. **41**, 533–568 (1992)
19. L.J. Lindsay, R.D. Mauldin Quantization dimension for conformal iterated function systems. Nonlinearity **15**, 189–199 (2002)
20. R.D. Mauldin, S.C. Williams, Hausdorff dimension Graph-directed constructions. Trans. AMS. Math. **309**, 811–829 (1988)
21. C. McMullen, The Hausdorff dimension of general Sierpiński carpetes. Nagoya Math. J. **96**, 1–9 (1984)
22. G. Pagès, A space quantization method for numerical integration. J. Comput. Appl. Math. **89**(1), 1–38 (1998)
23. G. Pagès, B. Wilbertz, Dual quantization for random walks with application to credit derivatives. J. Comput. Finance **16**, 33–60 (2012)
24. G. Pagès, B. Wilbertz, Optimal Delaunay and Voronoi quantization schemes for pricing American style options, in *Numerical Methods in Finance*. Springer Proc. Math., Vol. 12. (Springer, Heidelberg, 2012), pp. 171–213
25. Y. Peres, The self-affine carpetes of McMullen and Bedford have infinite Hausdorff measure. Math. Proc. Camb. Phil. Soc. **116**, 513–26 (1994)
26. K. Pözelberger, The quantization dimension of distributions. Math. Proc. Camb. Phil. Soc. **131**, 507–519 (2001)
27. M.K. Roychowdhury, Quantization dimension for some Moran measures. Proc. Am. Soc. **138**(11), 4045–4057 (2010)
28. M.K. Roychowdhury, Quantization dimension function and Gibbs measure associated with Moran set. J. Math. Anal. Appl. **373**(1), 73–82 (2011)
29. M.K. Roychowdhury, Quantization dimension and temperature function for recurrent self-similar measures. Chaos Solitons Fractals **44**(11), 947–953 (2011)
30. M.K. Roychowdhury, Quantization dimension estimate of probability measures on hyperbolic recurrent sets. Dyn. Syst. **29**(2), 225–238 (2014)

31. Z.Y. Wen, Moran sets and Moran classes. Chinese Sci. Bull. **46**, 1849–1856 (2001)
32. P.L. Zador, Development and evaluation of procedures for quantizing multivariate distributions. Ph.D. thesis, Stanford University (1964)
33. S. Zhu, The lower quantization coefficient of the F-conformal measure is positive. Nonlinear Anal. **69**(2), 448–455 (2008)
34. S. Zhu, Quantization dimension of probability measures supported on Cantor-like sets. J. Math. Anal. Appl. **338**(1), 742–750 (2008)
35. S. Zhu, Quantization dimension for condensation systems. Math. Z. **259**(1), 33–43 (2008)
36. S. Zhu, Quantization dimension for condensation systems. II. The geometric mean error. J. Math. Anal. Appl. **344**(1), 583–591 (2008)
37. S. Zhu, The quantization for self-conformal measures with respect to the geometric mean error. Nonlinearity **23**(11), 2849–2866 (2010)
38. S. Zhu, The quantization dimension for self-affine measures on general Sierpiński carpets. Monatsh. Math. **162**, 355–374 (2011)
39. S. Zhu, Asymptotic uniformity of the quantization error of self-similar measures. Math. Z. **267**(3–4), 915–929 (2011)
40. S. Zhu, On the upper and lower quantization coefficient for probability measures on multiscale Moran sets. Chaos, Solitons Fractals **45**, 1437–1443 (2012)
41. S. Zhu, A note on the quantization for probability measures with respect to the geometric mean error. Monatsh. Math. **167**(2), 295–304 (2012)
42. S. Zhu, Asymptotics of the geometric mean error in the quantization for product measures on Moran sets. J. Math. Anal. Appl. **403**(1), 252–261 (2013)
43. S. Zhu, A characterization of the optimal sets for self-similar measures with respect to the geometric mean error. Acta Math. Hungar. **138**(3), 201–225 (2013)
44. S. Zhu, Convergence order of the geometric mean errors for Markov-type measures. Chaos, Solitons Fractals **71**, 14–21 (2015)

Apollonian Circle Packings

Mark Pollicott

Abstract Circle packings are a particularly elegant and simple way to construct quite complicated and elaborate sets in the plane. One systematically constructs a countable family of tangent circles whose radii tend to zero. Although there are many problems in understanding all of the individual values of their radii, there is a particularly simple asymptotic formula for the radii of the circles, originally due to Kontorovich and Oh. In this partly expository note we will discuss the history of this problem, explain the asymptotic result and present an alternative approach.

Keywords Apollonian circle packings • Poincaré series • Transfer operators • Asymptotic formulae

Mathematics Subject Classification (2000). Primary 52C26, 37C30; Secondary 11K55, 37F35, 37D35

1 A Brief History of Apollonian Circles

Apollonius (c. 262–190 BC) was born in Perga (now in Turkey) and gave the names to various types of curves still used: ellipse, hyperbola and parabola. However, very little detail is known about his life and, although he wrote extensively on many topics, rather little of his work has survived (perhaps partly because it was considered too esoteric by his contemporaries). What has survived (partly in the form of translations into arabic) includes seven of his eight books on "conics". These include problems on tangencies of circles.

I am very grateful to Richard Sharp for many discussions on this approach and the details. I would also like to thank Christoph Bandt and the referee for many useful comments on the presentation.

M. Pollicott (✉)
Department of Mathematics, Zeeman Building, Warwick University, Coventry, CV4 7AL, UK
e-mail: masdbl@warwick.ac.uk

C. Bandt et al. (eds.), *Fractal Geometry and Stochastics V*, Progress in Probability 70, DOI 10.1007/978-3-319-18660-3_8

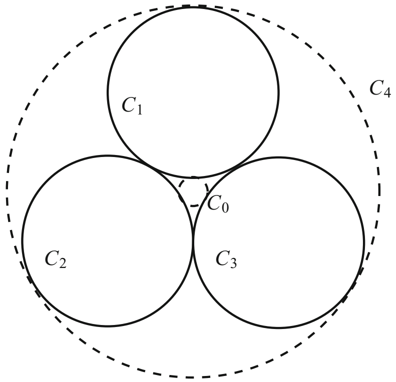


Fig. 1 The three initial circles C_1, C_2, C_3, and the two mutually tangent circles C_0 and C_4 guaranteed by Apollonius' theorem

The result of Apollonius which is of particular interest to us is the following.

Theorem 1.1 *Given three mutually tangent circles C_1, C_2, C_3 with disjoint interiors there are precisely two circles C_0, C_4 which are tangent to each of the original three.*

This result is illustrated in Fig. 1b. The proof is so easy and short that we include it.

Proof We can apply a Möbius transformation which takes a point of tangency between two of the initial circles to infinity. These two circles are then mapped to two parallel lines, and the third initial circle to a circle between, and just touching, these parallel lines. We can then construct the two new circles by translating the middle circle between the parallel lines and then transforming back. Since a Möbius transformation preserves circles and lines we are done. □

In 1643, **René Descartes** (1596–1650) wrote to Princess Elizabeth of Bohemia (1618–1680) stating a formula he had established on the radii a_1, a_2, a_3, a_4 of the tangent circles, and for which she independently provided a proof. The radii are related by the following formula.

Theorem 1.2 (Descartes-Princess Elizabeth) *Assume that the radii of the original circles are $a_1, a_2, a_3 > 0$ and the fourth mutually tangent circle has radius $a_4 > 0$ then*

$$2\left(\frac{1}{a_1^2}+\frac{1}{a_2^2}+\frac{1}{a_3^2}+\frac{1}{a_4^2}\right)=\left(\frac{1}{a_1}+\frac{1}{a_2}+\frac{1}{a_3}+\frac{1}{a_4}\right)^2.$$

A simple proof appears in the notes of Sarnak [16].

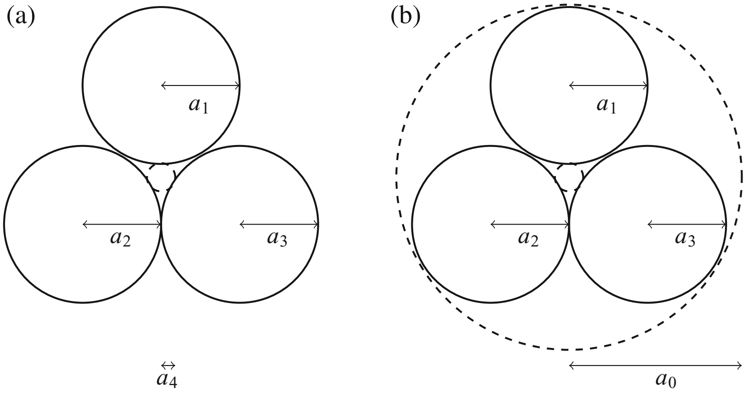


Fig. 2 (**a**) The initial circles with radii a_1, a_2, a_3 and smaller choice of mutually tangent circle with radius a_4; (**b**) The initial circles with radii a_1, a_2, a_3 and the larger choice of mutually tangent circle with radius a_0

Notation 1.3 *The formula also applies where the radius a_4 of the inner circle is replaced by the radius of the outer circle a_0. However, in this case we adopt the convention that $a_0 < 0$, where $|a_0| > 0$ is the radius of the circle C_0 (Fig. 2).*

Princess Elizabeth was a genuine princess by virtue of being the daughter of Queen Elizabeth (1596–1662) and King Frederick V of Bohemia (whose reign lasted a brief 1 year and 4 days). Queen Elizabeth of Bohemia was in turn the daughter of King James I of England (Fig. 3).

In 1605, King James was the target of an unsuccessful assassination plan (the "gunpowder plot" of Guy Fawkes and co-conspirators, celebrated in England annually on 5th November) and Queen Elizabeth of Bohemia would have become Queen of England (aged 9) had the plot succeeded.

In 1646, Elizabeth's brother Philip stabbed to death Monsieur L'Espinay, for flirting with their mother and sister. In the ensuing family rift, Elizabeth wrote to Queen Christina of Sweden for an audience and help reinstating her Father's lands, but Christina invited Descartes to Stockholm instead, which proved unfortunate for him since he promptly died of pneumonia. Finally, Elizabeth entered a convent in Germany for the last few years of her life, where she worked her way up to the top job of abbess.

The formula of Descartes was subsequently rediscovered by Frederick Soddy (1877–1956), which is the reason that the circles are sometimes called "Soddy circles". Frederick Soddy is more famous (outside of Mathematics) for having won the Nobel Prize for Chemistry in 1921, and having introduced the terms "isotopes" and "chain reaction". However, most relevant to us, he rediscovered the formula of Descartes and published it in the distinguished scientific journal *Nature* in the form of a poem [18]:

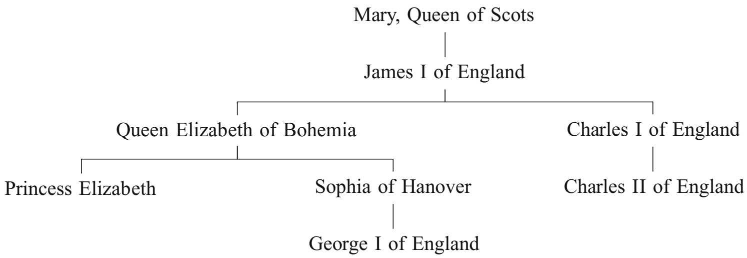


Fig. 3 The family tree of Princess Elizabeth. Her uncle, Charles I of England, was executed during the English revolution. Her nephew, George I, also became King of England and was the 6th Great-Grandfather of the present Queen

The kiss precise

For pairs of lips to kiss maybe
Involves no trigonometry.
'Tis not so when four circles kiss
Each one the other three.
To bring this off the four must be
As three in one or one in three.
If one in three, beyond a doubt
Each gets three kisses from without.
If three in one, then is that one
Thrice kissed internally.

Four circles to the kissing come.
The smaller are the benter.
The bend is just the inverse of
The distance from the center.
Though their intrigue left Euclid dumb
There's now no need for rule of thumb.
Since zero bend's a dead straight line
And concave bends have minus sign,
The sum of the squares of all four bends
Is half the square of their sum.

To spy out spherical affairs
An ocular surveyor
Might find the task laborious,
The sphere is much the gayer,
And now besides the pair of pairs
A fifth sphere in the kissing shares.
Yet, signs and zero as before,
For each to kiss the other four
The square of the sum of all five bends

Frederick Soddy (1877–1956)

2 Circle Counting

2.1 The Asymptotic Formulae

Starting from mutually tangent circles we can inscribe new circles inductively to arrive at what is known as an *Apollonian circle packing* consisting of infinitely many circles. We denote by $\mathcal{C}$ the set of such circles (Fig. 4).[1]

We can order these circles by (the reciprocal of) their radii, which we shall denote by a_n, for $n \geq 0$. It is easy to see that the sequence $(1/a_n)$ tends to infinity or, equivalently, the sequence of radii (a_n) tends to zero. This is because the total area of the disjoint disks enclosed by the circles $\sum_{n=1}^{\infty} \pi a_n^2$ which is in turn bounded by the area inside the outer circle. A natural question is then to ask: *How fast does the sequence* $(1/a_n)$ *grow, or, equivalently, how fast do the radii* (a_n) *tend to zero?* We begin with some notation.

Definition 2.1 Given, $T > 0$ we denote by $N(T)$ the finite number of circles with radii greater than $\frac{1}{T}$.

In particular, we see from our previous comments that $N(T) \to +\infty$ as $T \to 0$. A far stronger result is the following [5, 11].

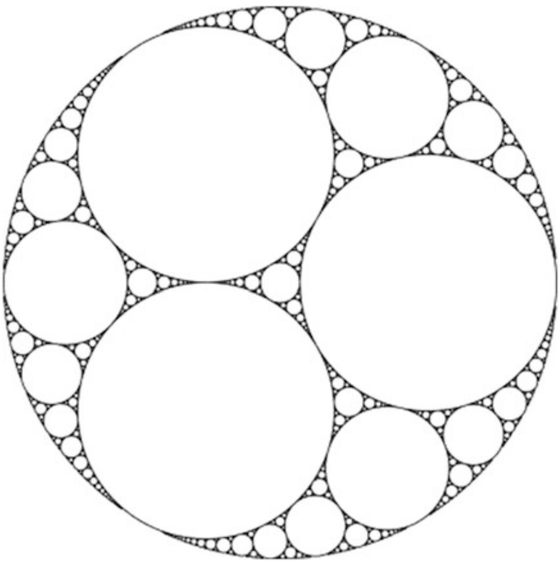


Fig. 4 An Apollonian circle packing consisting of infinitely many circles. The closure of their union is the Apollonian gasket denoted by Λ

[1]For other aspects of the rich theory of circle packings, we refer the reader to [19].

Theorem 2.2 (Kontorovich-Oh, 2009) *There exists $C > 0$ and $\delta > 1$ such that the number $N(T)$ is asymptotic to CT^δ as T tends to infinity, i.e.,*

$$\lim_{T\to\infty} \frac{N(T)}{T^\delta} = C.$$

It is the convention to write $N(T) \sim CT^\delta$ as $T \to \infty$ (Fig. 5).

We can illustrate Theorem 2.2 with two examples.

Example 1 Assume that we begin with four mutually tangent circles the reciprocals of whose radii are $a_0 = -\frac{1}{3}$, $a_1 = \frac{1}{5}$, $a_2 = \frac{1}{8}$ and $a_3 = \frac{1}{8}$. Using Theorem 1.2 we can compute the following monotone increasing sequence of reciprocal radii:

$$\left(\frac{1}{a_n}\right)_{n=1}^{\infty} = 5, 8, 8, 12, 12, 20, 20, 21, 29, 29, 32, 32, \cdots$$

We will return to this in Example 4 in the Appendix.

Example 2 Assume that we begin with four mutually tangent circles the reciprocals of whose radii are $a_0 = -\frac{1}{2}$, $a_1 = \frac{1}{3}$, $a_2 = \frac{1}{6}$ and $a_3 = \frac{1}{7}$. Using Theorem 1.2 we can compute the following monotone increasing sequence of reciprocal radii:

$$\left(\frac{1}{a_n}\right)_{n=1}^{\infty} = 3, 6, 7, 7, 10, 10, 15, 15, 19, 19, 22, 22, \cdots$$

We will return to this in Example 5 in the Appendix.

In the Appendix we also recall why the numbers in these sequences are all natural numbers.

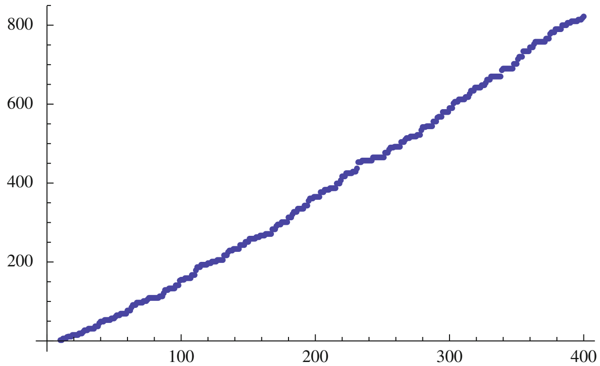


Fig. 5 A plot of $N(\epsilon)$ against $\frac{1}{\epsilon}$

2.2 The Exponent δ in Theorem 2.2

Of particular interest is the value of δ which controls the rate of growth of the radii. The next lemma provides an alternative characterisation of this number.

Lemma 2.3 *The value δ in Theorem 2.2 has the following alternative characterisation:*

$$\delta = \inf\left\{ t > 0 : \sum_{n=1}^{\infty} \frac{1}{a_n^t} < +\infty \right\}.$$

The expression for δ in Lemma 2.3 is usually called the *packing exponent*.

Notation 2.4 *We can denote by Λ the compact set given by the closure of the union of the circles in the Apollonian circle packing.*

This leads to a second useful alternative characterisation.

Lemma 2.5 *The value δ in Theorem 2.2 is equal to the Hausdorff Dimension $dim_H(\Lambda)$ of the limit set Λ.*

Remark 1 (The numerical value of δ) Unfortunately, there is no explicit expression for δ and it is rather difficult to estimate. The first rigorous estimates were due to Boyd [3] who, using the definition above, estimated $1{\cdot}300197 < \delta < 1{\cdot}314534$. A well known estimate is due to McMullen [10], who showed that $\delta = 1{\cdot}30568\ldots$

Perhaps a little surprisingly, the value of δ is independent of the particular Apollonian circle packing being considered, as is shown by the next lemma.

Lemma 2.6 *For different Apollonian circle packings exactly the same value of δ arises (independently of the initial choices a_0, a_1, a_2, a_3, a_4).*

Again the idea of the proof is so simple that we recall the idea so as to dispel any mystery.

Proof Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be any two Apollonian circle packings and let Λ_1 and Λ_2 be the associated Apollonian gaskets. By Lemma 2.5 it suffices to show that $\dim_H(\Lambda_1) = \dim_H(\Lambda_2)$. We can then deduce the independence of the value δ using the following well known result: If there exists a smooth bijection $T : \mathcal{C}_1 \to \mathcal{C}_2$ then the sets share the same Hausdorff Dimension. Let us identify the plane with $\mathbb{C}$. Then it is a simple exercise to show that there is a *Möbius transformation* $g : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ of the form

$$g(z) = \frac{az + b}{\overline{b}z + \overline{a}} \text{ and } a, b \in \mathbb{C} \text{ with } |a|^2 - |b|^2 = 1,$$

such that $T(\mathcal{C}_1) = \mathcal{C}_2$. In particular, this follows easily since Möbius transformations necessarily take circles to circles. □

3 Some Preliminaries for a Proof

We will describe a proof which differs from the original proof of Kontorovich-Oh and other subsequent proofs. This approach is more in the spirit of the classical proof of the Prime Number Theorem, except we use approximating Poincaré series in place of zeta functions.

3.1 An Analogy with the Prime Numbers

Purely for the purposes of motivation, we recall the classical Prime Number Theorem. Consider the prime numbers

$$2, 3, 5, 7, 11, 13, 17, 19, 23, \cdots$$

Let $\pi(x)$ denote the number of primes numbers between 1 and x. Since there are infinitely many primes, we see that $\pi(x) \to \infty$ as x tends to infinity. This again poses the natural question: *How does $\pi(x)$ grow as $x \to +\infty$?* The solution is the classical prime number Theorem [4].

Theorem 3.1 (Prime Number Theorem: Hadamard, de la Vallée Poussin (1896)) *There is a simple asymptotic formula $\pi(x) \sim \frac{x}{\log x}$ as $x \to +\infty$, i.e.,*

$$\lim_{x \to +\infty} \frac{\pi(x)}{\frac{x}{\log x}} = 1.$$

The essence of the proof of the Prime Number Theorem is to analyse the associated complex function, the *Riemann zeta function*, defined formally by

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}, \quad s \in \mathbb{C}.$$

The Riemann zeta function has the following important basic properties [4].

Lemma 3.2 *The Riemann zeta function $\zeta(s)$ converges to a well defined function for $Re(s) > 1$. Moreover:*

1. *For $Re(s) > 1$ we have that $\zeta(s)$ is analytic and non-zero;*
2. *There exists a small neighbourhood of each $1 + it$ with $t \neq 0$ on which $\zeta(s)$ has a non-zero analytic extension* [2]*; and*
3. *$\zeta(s)$ has a simple pole at $s = 1$.*

[2]The zeta function $\zeta(s)$ even has an analytic extension to $\mathbb{C} \setminus \{1\}$, but one does not need this to prove Theorem 3.1.

The result then follows by using a Tauberian theorem to convert this information on the domain of $\zeta(s)$ into information on prime numbers. For completeness, we recall the statement of the Ikehara-Wiener Tauberian Theorem [4].

Theorem 3.3 (Ikehara-Wiener Tauberian Theorem) *Assume that $\rho : \mathbb{R} \to \mathbb{R}$ is a monotone increasing function for which there exists $c > 0$, $\delta > 0$ such that the function*

$$F(s) := \int_0^\infty t^{-s} d\rho(t) - \frac{c}{s-\delta}$$

is analytic in a neighbourhood of $Re(s) \geq \delta$ then $\lim_{T\to+\infty} \frac{\rho(T)}{T^\delta} = c$.

Remark 2 The Prime Number Theorem easily follows from applying Theorem 3.3 to the auxiliary function $\rho(T) = \sum_{p^n \leq T} \log p$ and then relating the Stieltjes integral to $\zeta'(s)/\zeta(s)$. We refer the reader to [4] for further details of these now standard manipulations.

To adapt the proof of the Prime Number Theorem to the present setting, suggests considering a new complex function

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

where a_n are the radii of the circles in the Apollonian circle packing. In fact, it is more convenient to study a related function (a Poincaré series) and use an approximation argument to get the final result. However, to analyse such functions, we first introduce a dynamical ingredient.

3.2 An Iterated Function Scheme Viewpoint

Let us again identify the plane with the complex numbers $\mathbb{C}$, then we can introduce a transformation which preserves the circle packing $\mathcal{C}$. We want to define the "reflection" R in the circle $C = C(z_0, r)$ of radius r centered at z_0 (Fig. 6).

More precisely, let $z_0 \in \mathbb{C}$ and radius $r > 0$ then we associate a transformation

$$R : \mathbb{C} \setminus \{z_0\} \to \mathbb{C} \setminus \{z_0\}$$

$$R(z) = \frac{r^2(z-z_0)}{|z-z_0|^2} + z_0.$$

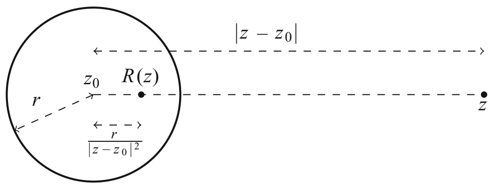


Fig. 6 Reflection in a circle

Rather than reflecting in the original Apollonian circles, we need to find four "dual circles" which we will reflect in. This point of view has a nice historical context. The original statement of the result was due to Philip Beecroft (1818–1862) who was a school teacher in Hyde, near Manchester, in England, and was the son of a miller and lived with his two elder sisters [1]. In his article he too had recovered Theorem 1.2.

Theorem 3.4 (Philip Beecroft, from "Lady's and Gentleman's diary" in 1842) *"If any four circles be described to touch each other mutually, another set of four circles of mutual contact may be described whose points of contact shall coincide with those of the first four."*

As in [5], we associate to the four initial Apollonian circles a new family of "dual" tangent circles (the dotted circles in Fig. 7). We can then consider the four associated reflections $R_i : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ in the four dual circles K_1, K_2, K_3, K_4 as shown in Fig. 7 ($i = 1, 2, 3, 4$).

The aim is to associate to the Apollonian circle packings complex functions, playing the rôle of the zeta function in number theory. These will be defined in terms of a family of contractions (i.e., an associated iterated function scheme) built out of the maps R_i on each of the four curvilinear triangles external to the initial four circles. For definiteness, let us fix the central curvilinear triangle Δ, whose sides are arcs from the circles C_1, C_2 and C_3 (with the other cases being similar) and let x_1, x_2, x_3 denote the vertices. We can consider the three natural linear fractional contractions $f_1, f_2, f_3 : \Delta \to \Delta$ defined by

$$f_i = R_4 \circ R_i, \quad i = 1, 2, 3,$$

each of which fixes the vertex x_i of Δ (Fig. 8). A simple calculation gives that:

- $|f_i'(z)| < 1$ for $z \in \Delta \setminus \{x_i\}$ for $i = 1, 2, 3$; and
- $|f_i'(x_i)| = 1$ for $i = 1, 2, 3$ (i.e., x_i is a parabolic point at the point of contact of K_4 with K_1, K_2 and K_3, respectively).

We recall the following explicit example from [7].

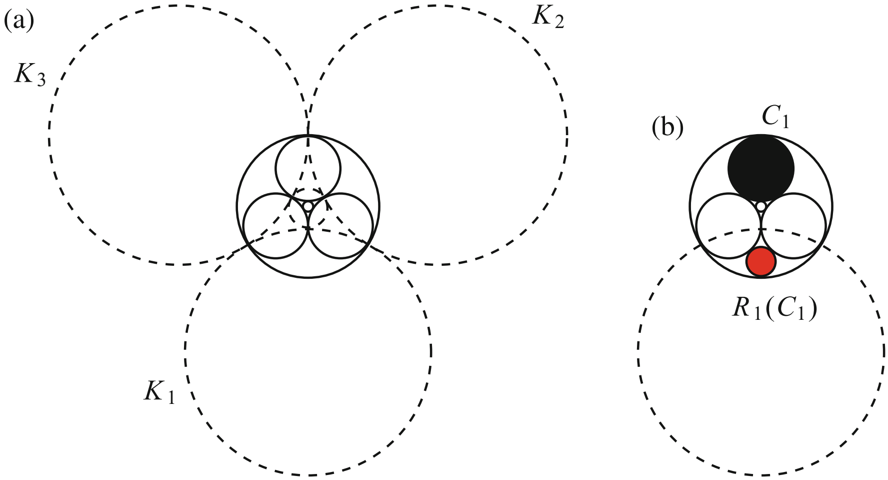


Fig. 7 (**a**) The four dual (*dotted*) circles K_1, K_2, K_3, K_4 associated to the original four Apollonian circles C_1, C_2, C_3, C_4; (b) The image of one of the original circles reflected in one of the dual circles begins the next generation of the circle packing

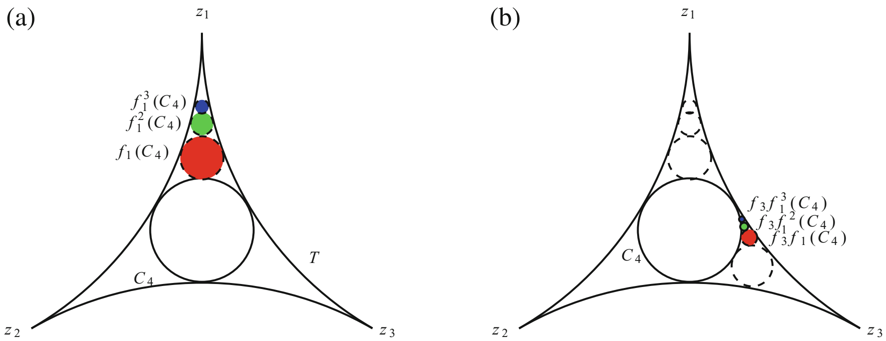


Fig. 8 (**a**) The central curvilinear triangle Δ and the images $f_1^n(C_4)$ of C_4 for $n = 1, 2, 3, \ldots$; (**b**) the images $f_3 f_1^n(C_4)$ of C_4 for $n = 1, 2, 3, \ldots$

Example 3 In the case of the Apollonian circle packing $\mathcal{C}$ with $a_0 = -1$ and $a_1 = a_2 = a_3$ we can explicitly write:

$$f_1(z) = \frac{az + b}{bz + a} \text{ where } a = -5\sqrt{\frac{4\sqrt{3} - 3}{78}} \text{ and } b = \sqrt{\frac{100\sqrt{3} - 153}{78}}$$

and $f_2(z) = e^{-2\pi i/3} f_1(e^{2\pi i/3} z)$ and $f_3(z) = e^{-2\pi i 2/3} f_1(e^{2\pi i 2/3} z)$.

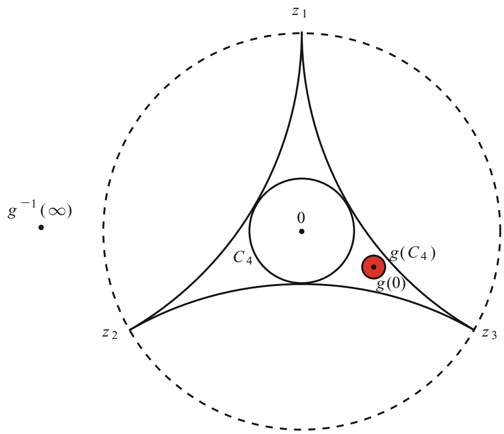


Fig. 9 The radius of $g(C_4)$ is related to the derivative $|g'(0)|$ by the value of $g^{-1}(\infty)$

In particular, one can easily check that:

1. For each $i = 1, 2, 3$ the iterates $f_i^k : \Delta \to \Delta$ $(k \geq 1)$ have the effect of mapping the central circle C_4 on to a sequences of circles $\{f_i^k(C_4)\}_{k=1}^{\infty}$ occurring in $\mathcal{C}$ leading into the vertex x_i (cf. Fig. 9a); and
2. Any sequence of compositions of these three maps can be naturally written in the form $\overline{f} := f_{i_k}^{n_k} \cdots f_{i_1}^{n_1} : \Delta \to \Delta$, for $n_1, \cdots, n_k \geq 1$ and $i_1, \cdots, i_k \in \{1, 2, 3\}$ with $i_l \neq i_{l+1}$ for $1 \leq l \leq k - 1$.

The relevance of these maps to our present study is that we see that we can rewrite

$$\xi(s) = \sum_{\overline{f}} \text{diam}(\overline{f}(C_0))^s,$$

at least for the contribution of circles in Δ, the other cases being similar, where the summation is over all such compositions $\overline{f} = f_{i_k}^{n_k} \cdots f_{i_1}^{n_1}$ in item 2 above.

3.3 Contracting Maps and Poincaré Series

The maps described above can be conveniently regrouped as follows:

$$\overline{f} := f_{i_k}^{n_k - 1} \circ (f_{i_k} \circ f_{i_{k-1}}^{n_k - 2}) \circ \cdots \circ (f_{i_{k-2}} \circ f_{i_2}^{n_2}) \circ (f_{i_2} \circ f_{i_1}^{n_1}). \tag{3.1}$$

The advantage of this presentation is that at least part of this expression is contracting, in the following sense (cf. [7]).

Lemma 3.5 (After Mauldin-Urbanski) *For the Apollonion circle packings we have that the maps* $\phi_j = \phi_j^{(i_j,n_j)} := f_{i_{j-1}} \circ f_{i_j}^{n_j} : \Delta \to \Delta$ *are uniformly contracting (i.e.,* $\sup_j \sup_{z\in T} |\phi_j'(z)| < 1$*).*

This is illustrated in Fig. 9b with $f_3^n f_1$, $n \geq 1$.

Unfortunately, considering only compositions of the uniform contractions ϕ_j leads only to some of the circles in the circle packing $\mathcal{C}$. The rest of the circles require the final application of the maps $f_{i_k}^{n_k-1}$ in (3.1), which therefore also needs addressing. Moreover, the counting function we will actually use is a more localized version, which allows us to approximate the counting function for circles by a counting function for derivatives – for which the associated complex functions are easier to analyse. In particular, we want to analyse the following related complex functions.

Definition 3.6 Given $z_0 \in \Delta$ and an allowed word $\underline{j} = (j_1, \cdots, j_N)$, with $j_r \neq j_{r+1}$ for $r = 1, \cdots, N-1$, we can associate a localised Poincaré function

$$\eta^{\underline{j}}(s) = \sum_{k=0}^{\infty} \sum_{\overline{\phi}} |(f_i^k \circ \overline{\phi} \circ \phi_{\underline{j}})'(z_0)|^s \tag{3.2}$$

where:

1. We first apply a fixed contraction $\phi_{\underline{j}} = \phi_{j_N} \circ \cdots \circ \phi_{j_1}$;
2. We next sum over all subsequent allowed hyperbolic compositions $\overline{\phi} := \phi_{i_n} \circ \cdots \circ \phi_{i_{N+1}} : \Delta \to \Delta$; and, finally,
3. We sum over the "parabolic tails" f_i^k (where i is associated to $\phi_{i_n} = f_i \circ f_l^n$, say).

The need to consider the contribution from different $\phi_{\underline{j}}$ is an artefact of our method of approximation in the proof.

Remark 3 Poincaré series are more familiar in the context of Kleinian groups Γ acting on three dimensional hyperbolic space and its boundary, the extended complex plane $\hat{\mathbb{C}}$. Our analysis applies to the Poincaré series of many such groups. In the particular case of classical Schottky groups the analysis is easier, since one can dispense with the parabolic tail (i.e., item 3 above).

As we will soon see, each such Poincaré series satisfies the hypotheses of Theorem 3.3, which allows us to estimate the corresponding counting function defined as follows.

Definition 3.7 We define an associated counting function

$$M^{\underline{j}}(T) = \#\{f_i^k \circ \overline{\phi} \circ \phi_{\underline{j}} : |(f_i^k \circ \overline{\phi} \circ \phi_{\underline{j}})'(z_0)| \leq T\} \text{ for } T > 0.$$

Let $\Sigma = \{(i_n)_{n=1}^{\infty} : i_n \neq i_{n+1} \text{ for } n \geq 0\}$ and consider the cylinder

$$[\underline{j}] = \{(i_n)_{n=1}^{\infty} \in \Sigma : i_r = j_r, \text{ for } 1 \leq r \leq N\}.$$

In particular, in the next section we will use the Poincaré series to deduce the following.

Proposition 3.8 *There exists $C > 0$ and a measure μ on Σ such that $M^{\underline{j}}(T) \sim C\mu([\underline{j}])T^{\delta}$ as $T \to +\infty$, i.e.,*

$$\lim_{T\to+\infty} \frac{M^{\underline{j}}(T)}{T^{\delta}} = 1.$$

There may be some circles whose radii we don't seem to capture with this coding, but their contribution doesn't effect the basic asymptotics.

4 The Proof of Theorem 2.2

To complete the proof of Theorem 2.2 we need to complete the proof of Proposition 3.8 (in Sect. 4.1 below) and then perform the approximation of the counting functions for circles by those for derivatives (in Sect. 4.2 below).

4.1 Extending the Poincaré Series

By the chain rule we can write

$$(f_i^k \circ \overline{\phi} \circ \phi_{\underline{j}})'(z_0) = (f_i^k)'(\overline{\phi} \circ \phi_{\underline{j}} z_0)\overline{\phi}'(\phi_{\underline{j}} z_0)\phi_{\underline{j}}'(z_0)$$

and, in particular, we can now rewrite the expression for $\eta^{\underline{j}}(s)$ in (3.2) as:

$$\begin{aligned} \eta^{\underline{j}}(s) &= |\phi_{\underline{j}}'(z_0)|^s \sum_{n=0}^{\infty} \sum_{|\overline{\phi}|=n} \sum_{l=0}^{\infty} (f_i^k)'(\overline{\phi} \circ \phi_{\underline{j}} z_0)\overline{\phi}'(\phi_{\underline{j}} z_0) \\ &= |\phi_{\underline{j}}'(z_0)|^s \sum_{n=0}^{\infty} \sum_{|\overline{\phi}|=n} |\overline{\phi}'(z_0)|^s h_s(\overline{\phi}(z_0)) \end{aligned} \tag{4.1}$$

where the function $h_s : \Delta \to \mathbb{C}$ is defined by the summation

$$h_s(z) := \sum_{l=0}^{\infty} |(f_i^l)'(z)|^s \in C^1(\Delta)$$

is analytic in s. In particular, we see from the following lemma that $h_s(z)$ converges to a well defined function for $Re(s) > \frac{1}{2}$.

Lemma 4.1 *We can estimate* $||(f_i^l)'|_\Delta||_\infty = O(l^{-2})$.

We recall the simple proof (cf. [8]).

Proof By a linear fractional change of coordinates (mapping the vertex of Δ to infinity) the map f_i becomes a translation. Transforming this back to convenient coordinates we can write, say,

$$f_i^l(z) = \frac{(\sqrt{3}-l)z + l}{(-lz + l + \sqrt{3})}.$$

From this we see that

$$|(f_i^l)'(z)| = \frac{1}{|-lz + l + \sqrt{3}|^2}$$

and the required estimate follows. □

The Poincaré series have the useful feature that they can be expressed simply in terms of linear operators on appropriate Banach spaces of functions.

Definition 4.2 Let $C^1(\Delta)$ be the Banach space of C^1 functions on Δ. We can consider the transfer operators $\mathcal{L}_s : C^1(\Delta) \to C^1(\Delta)$ ($s \in \mathbb{C}$) given by

$$\mathcal{L}_s w(x) = \sum_l |\phi_l'(x)|^s w(\phi_l x)$$

where $w \in C^1(\Delta)$. This converges provided $Re(s) > \frac{1}{2}$.

We are actually spoilt for choice of Banach spaces. Although the continuous functions $C^0(\Delta)$ is too large a space for our purposes, we could also work with Hölder continuous functions or suitable analytic functions (on some neighbourhood of the complexification of Δ thought of as a subset of $\mathbb{R}^2$). The choice of $C^1(\Delta)$ is perhaps the more familiar.

The approach in the rest of this subsection is now relatively well known (cf. [6, 8, 13], for example) and is a variant on the symbolic approach to Poincaré series and the hyperbolic circle problem [14, 15]. Recall that $\delta > 0$ is the exponent in Theorem 2.2.

Lemma 4.3 *The operators are well defined provided* $Re(s) > \frac{1}{2}$. *Moreover, for* $Re(s) > \delta$ *we have that the spectral radius satisfies*

$$\rho(\mathcal{L}_s) := \limsup_{n\to+\infty} \|\mathcal{L}_s^n\|^{\frac{1}{n}} < 1.$$

In particular, we see from the definition of $\mathcal{L}_s$ that we can write

$$\mathcal{L}_s^n w(z) = \sum_{\overline{\phi}} |\overline{\phi}'(z)|^s w(\overline{\phi} z), \text{ for } n \geq 2,$$

where the summation is over allowed compositions of contractions $\overline{\phi} = \phi_{i_n} \circ \cdots \circ \phi_{i_1}$. We can now rewrite the expression for the Poincaré series in (4.1) more concisely as

$$\eta^{\underline{j}}(s) = |\phi_{\underline{j}}'(z_0)|^s \sum_{n=0}^{\infty} \mathcal{L}_s^n h_s(\overline{\phi}_{\underline{j}} z_0).$$

In order to construct the required extension of $\eta^{\underline{j}}(s)$, we recall the following simple lemma improving on the result in Lemma 4.3.

Lemma 4.4 *Let $Re(s) = \delta$. Then*

1. *For $s = \delta + it$ with $t \neq 0$ we have that the spectral radius satisfies $\rho(\mathcal{L}_s) < 1$; and*
2. *For $s = \delta$ we can write $\mathcal{L}_\delta = Q + U$ where*
   (a) *Q is a (one dimensional) eigenprojection with $QU = UQ = 0$, $Q^2 = Q$, and*
   (b) *And* $\limsup_{n \to +\infty} \|U^n\|^{1/n} < 1$.

Remark 4 The spectral properties of $\mathcal{L}_s$ can be seen when the operator acts on C^1 functions. Alternatively, we could have looked at bounded analytic functions on a small enough neighbourhood $T \subset U \subset \mathbb{C}^2$ in the complexification (cf. [6]).

We can now deduce almost immediately from Lemmas 4.3 and 4.4 the following corollary for this Poincaré series.

Corollary 4.5 *The Poincaré series $\eta^{\underline{j}}(s)$ converges to a well defined function on $Re(s) > \delta$. Moreover,*

1. *For $Re(s) > \delta$ we have that $\eta^{\underline{j}}(s)$ is analytic;*
2. *There exists a small neighbourhood of each $\delta + it$ with $t \neq 0$ on which $\eta^{\underline{j}}(s)$ has an analytic extension; and*
3. *$\eta^{\underline{j}}(s)$ has a simple pole at $s = \delta$.*

Remark 5 In fact, we can deduce a little more which, if a little technical looking, is needed in the approximation argument below. In particular, we can also show that the simple pole for $\eta^{\underline{j}}(s)$ at $s = \delta$ has a residue of the form

$$C_{\underline{j}} := \frac{|(\phi_{\underline{j}})'(x_0)|^\delta \mu(h_s)}{\lambda'(\delta)}$$

where:

(i) $\lambda(t)$ is an isolated eigenvalue equal to the spectral radius of $\mathcal{L}_t$ ($t \in \mathbb{R}$); and
(ii) $Q(h) = \mu(h)k$ where k is an associated eigenfunction, i.e., $\mathcal{L}_1 k = k$.

If we now write

$$\eta^{\underline{j}}(s) = \int_1^\infty t^{-s} dN^{\underline{j}}(t)$$

then comparing Corollary 4.5 with Theorem 3.3 gives the asymptotic formula for $N^{j}_{\underline{l}}(T)$ in Proposition 3.8.

Let us now move on to the final step in the proof of Theorem 2.2.

4.2 The Approximation Argument

We can now approximate the radii $\mathrm{rad}(g(C_4))$ of the circle $g(C_4)$ by suitably scaled values of $1/|g'(x_0)|$, where $g = f_i^l \circ \overline{\phi} \circ \phi_{\underline{j}}$. Without loss of generality we can choose coordinates in $\mathbb{C}$ so that C_4 is the unit circle.

As a prelude to this we consider some simple geometric estimates on the sizes of the images of circles.

Lemma 4.6 *If $g(z) = (az+b)/(cz+d)$, with $ad - bc = 1$ and $a, b, c, d \in \mathbb{C}$, then the radius of the image circle $C = g(C_4)$ is equal to*

$$\frac{1}{||c|^2 - |d|^2|} = \frac{|g'(0)|}{||\frac{c}{d}|^2 - 1|}$$

The proof is a reassuringly elementary exercise:

Proof For the first part, we see that the image circle $g(C_0)$ has centre $z_c = (a\overline{c} - b\overline{d})/(|c|^2 - |d|^2)$ and radius $1/(||c|^2 - |d|^2|)$ since we can check that for $e^{i\theta} \in C_4 = \{z \in \mathbb{C} : |z| = 1\}$:

$$|g(e^{i\theta}) - z_c| = \left| \frac{ae^{i\theta} + b}{ce^{i\theta} + d} - \frac{a\overline{c} - b\overline{c}}{|c|^2 - |d|^2} \right| = \frac{1}{||c|^2 - |d|^2|}.$$

We then observe that $|g'(z)| = |cz + d|^{-2}$ and thus $|g'(0)| = |d|^{-2}$. Thus by the above we see that the radius of the image circle C is:

$$\mathrm{rad}(C) = \frac{1}{||c|^2 - |d|^2|} = \frac{|g'(0)|}{\left||\frac{c}{d}|^2 - 1\right|}.$$

as claimed. □

We can write $g^{-1}(z) = (dz - b)/(-cz + a)$ and thus $g^{-1}(\infty) = d/c$.

Finally, we come to the crux of the approximation argument. The essential idea is to approximate the (technically more convenient) weighting of elements g by $|g'(z_0)|$, with a weighting by the more geometric weighting by reciprocals of the radii $\mathrm{rad}(g(C_4))$. One simple approach is as follows. We are taking $z_0 = 0$, for definiteness, and then we want to use Proposition 3.8 to localise the counting to regions where

$$\frac{|g'(0)|}{\mathrm{rad}(g(C_4))} = \left| \left| \frac{c}{d} \right|^2 - 1 \right|$$

is close to constant, using Lemma 4.6. Given an allowed string $(j_1, \cdots, j_N)$ we can write

$$\begin{aligned} g^{-1} &= \left((R_4 \circ R_{j_k})^{n_k} \circ \cdots \circ (R_4 \circ R_{j_{N+1}})^{n_{N+1}} \circ (R_4 \circ R_{j_N})^{n_N} \circ \cdots \circ (R_4 \circ R_{j_1})^{n_1}\right)^{-1} \\ &= (R_{j_1} \circ R_4)^{n_1} \circ \cdots \circ (R_{j_N} \circ R_4)^{n_N} \circ (R_{j_{N+1}} \circ R_4)^{n_{N+1}} \circ \cdots \circ (R_{j_k} \circ R_4)^{n_k} \\ &= \overline{f}_{j_k}^{n_k} \circ \cdots \circ \overline{f}_{j_1}^{n_1}, \end{aligned}$$

where we denote $\overline{f}_j := R_j \circ R_4$ $(j = 1, 2, 3)$ acting on the complement of the disk containing Δ (i.e., the dotted circle in Fig. 10). In particular, given $\eta > 0$, we can choose N sufficiently large such that for each $|\underline{j}| = N$ we can choose $K_{\underline{j}}$ such that for $g = f_i^l \circ \overline{\phi} \circ \phi_{\underline{j}}$:

$$K_{\underline{j}} - \eta \leq \frac{|g'(0)|}{\mathrm{rad}(g(C_4))} \leq K_{\underline{j}} + \eta. \tag{4.2}$$

We can define a local version of $N(T)$, which is useful to compare with $M^{\underline{j}}(T)$.

Definition 4.7 We define a restricted counting function

$$N^{\underline{j}}(T) = \{g : \mathrm{rad}(g(C_4)) \leq T\},$$

for $T > 0$.

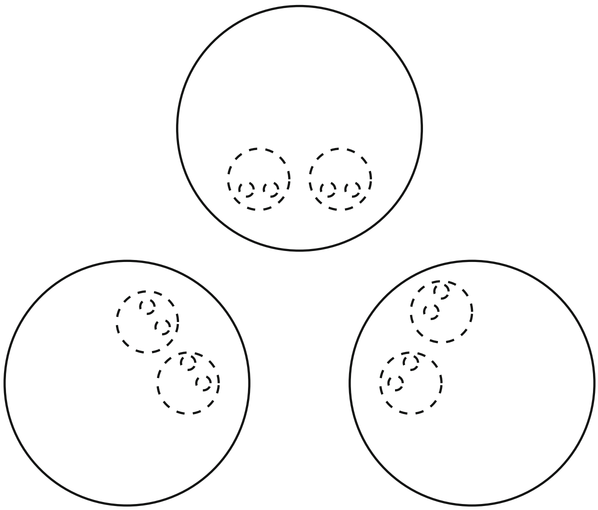


Fig. 10 Sequences of circles generated by reflections in disjoint circles. The three initial circles are represented by *solid lines* and the first two generations of circles generated by reflections are represented by *dashed lines*

Using (4.2) we can write

$$M^{\underline{j}}\left(\frac{T}{K_{\underline{j}}+\eta}\right) \le N^{\underline{j}}(T) \le M^{\underline{j}}\left(\frac{T}{K_{\underline{j}}-\eta}\right).$$

and observe that $N(T) = \sum_{|\underline{j}|=N} N^{\underline{j}}(T)$. Using the asymptotic formula from Proposition 3.8 and summing over allowed strings $\underline{j}$ of length N, we have that

$$C \sum_{|\underline{j}|=N} \frac{\mu([\underline{j}])}{(K_{\underline{j}}+\eta)^{\delta}} \le \liminf_{T\to\infty} \frac{N(T)}{T^{\delta}} \le \limsup_{T\to\infty} \frac{N(T)}{T^{\delta}} \le C \sum_{|\underline{j}|=N} \frac{\mu([\underline{j}])}{(K_{\underline{j}}-\eta)^{\delta}}$$

Letting $N \to +\infty$ (and thus $\epsilon \to 0$) gives the result in Theorem 2.2 with

$$K = \lim_{N\to+\infty} C \sum_{|\underline{j}|=N} \frac{\mu([\underline{j}])}{(K_{\underline{j}})^{\delta}}.$$

Remark 6 The existence of the limit, and its value K, can be understood in terms of an integral related to the natural measure μ on $\mathcal{C}$. A modified argument leads to an equidistribution result (expressed in terms of the measure μ, of course).

4.3 Generalizations

The approach to counting circles is more analytical than geometrical, and thus is somewhat oblivious to the specific setting of circle packings. In particular, the same method of proof works in a number of related settings where we ask for the radii of circles which are images under a suitable Kleinian group. For example:

1. Other circle packings for which the circles can be generated by the image of circles under reflections;
2. The radii of the images $g(C)$ of a circle C, where $\Gamma \subset SL(2,\mathbb{C})$ is a Schottky group (i.e., a convex cocompact Kleinian group generated by reflections in a finite number of circles with disjoint interiors);
3. The radii of the images $g(C)$ of a circle C, where $\Gamma \subset SL(2,\mathbb{C})$ is a quasi-Fuchsian group.

For more details of such examples, we refer the reader to [9].

The same basic method can also be used to prove other more subtle statistical properties of the radii of the circles.

Appendix: The Case of Reciprocal Integer Circles

The following is an interesting corollary to Descartes' Theorem.

Corollary 4.8 *If* $\frac{1}{a_0}, \frac{1}{a_1}, \frac{1}{a_2}, \frac{1}{a_3} \in \mathbb{Z}$ *then* $\frac{1}{a_4} \in \mathbb{Z}$.

Proof In particular, this is a quadratic polynomial in $\frac{1}{a_4} > 0$, so given the radii of the initial circles a_1, a_2, a_3 we have two possible solutions

$$\frac{1}{a_1} + \frac{1}{a_2} + \frac{1}{a_3} \pm 2\sqrt{\frac{1}{a_1 a_2} + \frac{1}{a_2 a_3} + \frac{1}{a_3 a_4}}.$$

and we denote these $\frac{1}{a_4} > 0$ (and $\frac{1}{a_0} < 0$). We use the convention that the smaller inner circle has radius $a_4 > 0$ and the larger outer circle has a negative "radius" a_4 (meaning its actually radius is $|a_4| > 0$ and the negative sign just tells us it is the outer circle). Adding these two solutions gives:

$$\frac{1}{a_0} + \frac{1}{a_4} = 2\left(\frac{1}{a_1} + \frac{1}{a_2} + \frac{1}{a_3}\right)$$

from which we easily deduce the result. □

Proceeding inductively, for any subsequent configuration of four circles with radii $a_n, a_{n+1}, a_{n+2}, a_{n+3}$, for $n \geq 0$, we can similarly write

$$\frac{1}{a_{n+4}} = 2\left(\frac{1}{a_{n+1}} + \frac{1}{a_{n+2}} + \frac{1}{a_{n+3}}\right) - \frac{1}{a_n}.$$

Proceeding inductively, then one gets infinitely many circles. Moreover, if the reciprocals of the initial four circles are integers then we easily see that this is true for all subsequence circles.

Corollary 4.9 *If the four initial Apollonian circles have that their radii* a_0, a_1, a_2, a_3 *are reciprocals of integers then all of the circles in* $\mathcal{C}$ *have that the reciprocals of their radii* a_n, $n \geq 4$, *are integers.*

Example 4 Let us consider the example starting with $a_0 = -\frac{1}{3}$, $a_1 = \frac{1}{5}$, $a_2 = \frac{1}{8}$, and $a_3 = \frac{1}{8}$. In Fig. 11 below we illustrate the iterative process of inscribing circles into each curved triangle formed by three previously constructed tangent circle and write $\frac{1}{a_n}$ inside the corresponding circle of radius a_n.

Example 5 Let us also consider the example with $a_0 = -\frac{1}{2}$, $a_1 = \frac{1}{3}$, $a_2 = \frac{1}{6}$, $a_3 = \frac{1}{7}$. In Fig. 12 below we illustrate the iterative process of inscribing circles into each curved triangle formed by three previously constructed tangent circle and write $\frac{1}{a_n}$ inside the corresponding circle of radius a_n.

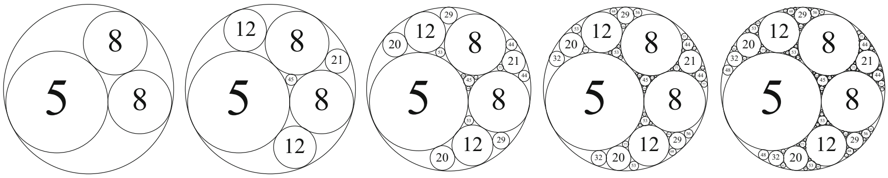


Fig. 11 We iteratively inscribe additional circles starting with circles of radii $a_0 = -\frac{1}{3}$, $a_1 = \frac{1}{5}$, $a_2 = \frac{1}{8}$, $a_3 = \frac{1}{8}$

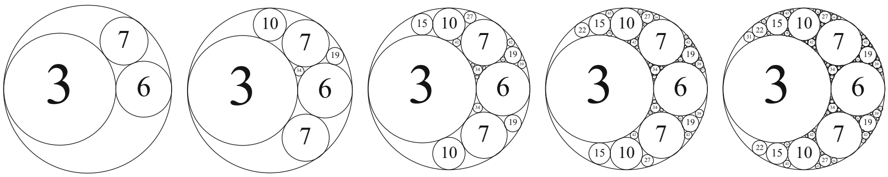


Fig. 12 We iteratively inscribe additional circles starting with circles with radii $a_0 = -\frac{1}{2}$, $a_1 = \frac{1}{3}$, $a_2 = \frac{1}{6}$, $a_3 = \frac{1}{7}$

Remark 7 An easy consequence of the fact $\delta > 1$ is that then $\frac{1}{a_n} \in \mathbb{N}$ some value must necessarily have high multiplicity (since we need to fit approximately $C\epsilon^{-\delta}$ inverse diameters into the first $[\epsilon^{-1}]$ natural numbers and the "pigeonhole principle" applies). In subsequent work, Oh-Shah showed that similar results are true for other sorts of circle packing [12]. Oh-Shah also gave an alternative approach to the original proof of Kontorovich-Oh using ideas of Roblin.

Remark 8 Another question we might ask is: *It we remove the repetitions in the sequence* (a_n) *then how many distinct diameters are greater than* ϵ*?* The following result was proved by Bourgain and Fuchs [2]: There exists $C > 0$ such that

$$\#\{\text{distinct diameters } a_n : a_n \geq \epsilon\} \geq \frac{C}{\epsilon}$$

for all sufficiently large ϵ.s Previously, Sarnak [17] had proved the slightly weaker result that there exists $C > 0$ such that

$$\#\{\text{distinct diameters } a_n : a_n \geq \epsilon\} \geq \frac{C}{\epsilon\sqrt{\log \epsilon}}.$$

References

1. P. Beecroft, Properties of circles in mutual contact. Lady's Gentleman's Diary **139**, 91–96 (1841)
2. J. Bourgain, E. Fuchs, A proof of the positive density conjecture for integer Apollonian circle packings. J. Am. Math. Soc. **24**, 945–967 (2011)
3. D. Boyd, The sequence of radii of the Apollonian packing. Math. Comput. **39**, 249–254 (1982)
4. W. Ellison, F. Ellison, *Prime Numbers* (Wiley, New York, 1985)
5. A. Kontorovich, H. Oh, Apollonian circle packings and closed horospheres on hyperbolic 3-manifolds. J. Am. Math. Soc. **24**, 603–648 (2011)
6. D. Mayer, Approach to equilibrium for locally expanding maps in $\mathbb{R}^k$. Commun. Math. Phys. **95**(1), 1–15 (1984)
7. R. Mauldin, M. Urbanski, Dimension and measures for a curvilinear Sierpinski gasket or Apollonian packing. Adv. Math. **136**, 26–38 (1998)
8. R. Mauldin, M. Urbanski, *Graph Directed Markov Systems, Geometry and Dynamics of Limit Sets*. Cambridge Tracts in Mathematics, vol. 148 (Cambridge University Press, Cambridge, 2003)
9. D. Mumford, C. Series, D. Wright, *Indra's Pearls* (Cambridge University Press, Cambridge, 2002)
10. C. McMullen, Hausdorff dimension and conformal dynamics. III. Computation of dimension. Am. J. Math. **120**, 691–721 (1998)
11. H. Oh, Dynamics on geometrically finite hyperbolic manifolds with applications to Apollonian circle packings and beyond, in *Proceedings of the International Congress of Mathematicians*, vol. III (Hindustan Book Agency, New Delhi, 2010), pp. 1308–1331
12. H. Oh, N. Shah, The asymptotic distribution of circles in the orbits of Kleinian groups. Invent. Math. **187**(1), 1–35 (2012)
13. W. Parry, M. Pollicott, *Zeta Functions and Closed Orbits for Hyperbolic Systems*. Asterisque (Soc. Math. Fr.) **187–188**, 1–268 (1990)
14. M. Pollicott, R. Sharp, Orbit counting for some discrete groups acting on simply connected manifolds with negative curvature. Invent. Math. **117**, 275–302 (1994)
15. M. Pollicott, R. Sharp, Comparison theorems in hyperbolic geometry. Trans. Am. Math. Soc. **350**, 473–499 (1998)
16. P. Sarnak, Integral Apollonian packings. Am. Math. Mon. **118**, 291–306 (2011)
17. P. Sarnak, Letter to Lagarias (2008). http://www.math.princeton.edu/sarnak
18. F. Soddy, The kiss precise. Nature **137**, 1021 (1936)
19. K. Stephenson, *Introduction to Circle Packing: The Theory of Discrete Analytic Functions* (Cambridge University Press, Cambridge, 2005)

Entropy of Lyapunov-Optimizing Measures of Some Matrix Cocycles

Michał Rams

Abstract This is an extended version of my talk at the Fractal Geometry and Stochastic V conference in Tabarz. It is based on my joint paper (Bochi and Rams, The entropy of Lyapunov-optimizing measures of some metric cocycles, preprint) with Jairo Bochi (PUC Santiago). Compared with the paper, I'll skip some details of some proofs, but I'll try to explain the main idea of our approach.

Keywords Joint spectral radius • Ergodic optimization • Barabanov norm • Noncommutative multifractal formalism

Mathematics Subject Classification (1991). Primary 15B48 Secondary 37H15, 37D30, 93C30

1 Setting

The object we study is seemingly very simple. We are given a finite family of 2×2 matrices $A_1, \ldots, A_k \in GL(2, \mathbb{R})$. For any sequence $\omega \in \{1, \ldots, k\}^{\mathbb{N}}$ we write $A_n(\omega) = A_{\omega_{n-1}} \cdot \ldots \cdot A_{\omega_0}$ and consider the *Lyapunov exponent*

$$\lambda(\omega) = \lim_{n\to\infty} \frac{1}{n} \log |A_n(\omega)|, \tag{1.1}$$

whenever it exists. The maximum and minimum values λ^+, λ^- attained by the Lyapunov exponent are called the *joint spectral radius* and *joint spectral subradius*, respectively; those notions play a significant role in control theory, see for example [11] and references therein.

The same object appears naturally in dynamical systems as well; let us explain the relation. Let us start from the main object studied in the area of multifractal

M. Rams (✉)
Institute of Mathematics, Polish Academy of Sciences, ul. Śniadeckich 8, 00-956 Warszawa, Poland
e-mail: rams@impan.pl

C. Bandt et al. (eds.), *Fractal Geometry and Stochastics V*, Progress in Probability 70, DOI 10.1007/978-3-319-18660-3_9

formalism: the Birkhoff average. Let $T : X \to X$ be a topological dynamical system (a continuous map of a compact space into itself) and let $\Phi : X \to \mathbb{R}^+$ be a continuous function. We consider the *cocycle* $\hat{T} : (X \times \mathbb{R}^+) \to (X \times \mathbb{R}^+)$ given by the formula

$$\tilde{T}(x, r) = (T(x), r \cdot \Phi(x)).$$

The value

$$\lambda(x) = \lim_{n\to\infty} \frac{1}{n} \log \left(\Phi(x) \cdot \Phi(Tx) \cdot \ldots \cdot \Phi(T^{n-1}x)\right)$$

(whenever it is defined) is called the *exponential rate of growth* in the fiber $\{x\}$, or the *Birkhoff average* of the potential $\log \Phi$ at the point x.

Let us now consider a natural generalization of this object: *noncommutative* Birkhoff averages. That is, we replace $\mathbb{R}^+$ by some noncommutative group, and we calculate the fiber rate of growth of the corresponding cocycle using some appropriate norm. In our case, the base dynamics is the full shift on k symbols, the fiber action is given by the group $GL(2, \mathbb{R})$ and the norm is the usual matrix norm:

$$\tilde{T}(\omega, M) = (\sigma\omega, A(\omega) \cdot M)$$

(where A is a 2×2 matrix-valued potential), so

$$\lambda(\omega) = \lim_{n\to\infty} \frac{1}{n} \log |A(\sigma^{n-1}\omega) \cdot \ldots \cdot A(\sigma\omega) \cdot A(\omega)|. \tag{1.2}$$

This system is quite complicated, so let us consider the special case: the *one-step* cocycle, that is, let $A(\omega)$ depend only on ω_0. This takes us exactly to the situation we considered in the beginning: denoting by A_ℓ the value of A on $\{\omega; \omega_0 = \ell\}$, (1.2) reduces to (1.1).

2 Domination

It turns out to be difficult to describe the product of matrices, in particular, the norm of such a product can strongly depend on the order in which we multiply the matrices. For this reason the usual dynamical approach is to forget about the geometry of matrix product, and use only the subadditivity property of the (logarithm of) norm. The theory of *subadditive thermodynamical formalism* has recently developed strongly, let us just mention the book [3] and the papers [9, 10].

We will apply an alternative approach, coming from the paper [6]. That is, instead of considering a product of matrices and asking how fast the norm grows, we will multiply this product by a given vector, and will ask how fast the length of the vector grows:

$$\lambda(\omega, v) = \lim_{n\to\infty} \frac{1}{n} \log |A_n(\omega) \cdot v| .$$

The main difference is that we can write

$$\log |A_n(\omega) \cdot v| = \sum_{\ell=0}^{n-1} \log \left| A_{\omega_\ell} \frac{A_\ell(\omega)v}{|A_\ell(\omega)v|} \right| .$$

That is, we replace a noncommutative cocycle over a simple dynamical system (full shift) by a commutative cocycle but over a considerably more complicated system (action of matrices $\{A_i\}$ on $\mathbb{P}^1$). However, we need to explain why the growth rate of the length of a vector is related to the growth rate of the matrix norm in our original problem.

It will be more convenient for us to work with cocycles over the full shift on bi-infinite sequences (Σ, σ), where $\Sigma = \{1, \ldots, k\}^{\mathbb{Z}}$ and σ is the usual left shift. Naturally, the Lyapunov exponent $\lambda(\omega)$ can be defined on this space as well; it will only depend on the positive coordinates $\omega_+ = \{\omega_i, i \geq 0\}$. We will distinguish between the action of a matrix on $\mathbb{R}^2_* = \mathbb{R}^2 \setminus \{0, 0\}$ and the action on $\mathbb{P}^1$ by the following notation: when we have $A : \mathbb{R}^2_* \to \mathbb{R}^2_*$, we write $A' : \mathbb{P}^1 \to \mathbb{P}^1$. Similarly, if M is a union of a family of lines in $\mathbb{R}^2_*$ passing through the origin, we denote by M' the corresponding subset of $\mathbb{P}^1$.

We say that the 2×2 matrix cocycle is *dominated* (or *exponentially separated*) if for each $\omega \in \Sigma$ we are given a splitting of $\mathbb{R}^2$ as the sum of two one-dimensional subspaces $e_1(\omega)$, $e_2(\omega)$ such that the following properties hold:

- Equivariance:

$$A(\omega)(e_i(\omega)) = e_i(\sigma\omega) \quad \text{for all } \omega \in \Sigma \text{ and } i \in \{1, 2\}; \tag{2.1}$$

- Dominance: there are constants $c > 0$ and $\delta > 0$ such that

$$\frac{|A^{(n)}(\omega)e_1(\omega)|}{|A^{(n)}(\omega)e_2(\omega)|} \geq ce^{\delta n} \quad \text{for all } \omega \in \Sigma \text{ and } n \geq 1. \tag{2.2}$$

This definition works for general cocycles, in our case there exists another, equivalent, definition. We define the *standard positive cone* in $\mathbb{R}^2_* := \mathbb{R}^2 \setminus \{0\}$ as

$$C_+ = \{(x, y) \in \mathbb{R}^2_*;\ xy \geq 0\}.$$

A *cone* in $\mathbb{R}^2_*$ is an image of C_+ by a linear isomorphism. A *multicone* in $\mathbb{R}^2_*$ is a disjoint union of finitely many cones. It was proved in [1, 4] that the one-step cocycle generated by $\{A_1, \ldots, A_k\}$ is dominated if and only if it has a *forward-invariant multicone*, that is, when there exists a multicone M such that its image $\bigcup_i A_i(M)$ is contained in the interior of M.

We can choose on M' a generalization of the Hilbert metric, that is a bounded metric d (depending on $\{A_1, \ldots, A_k\}$) in which all the maps A'_i are uniformly contracting:

Lemma 2.1 *Let $\{A_1, \ldots, A_k\}$ be a dominated cocycle with forward-invariant multicone M. Then there exists a metric d on M' and constants $c_0 > 1, 0 < \tau < 1$ such that for $v', w' \in M'$ we have*

$$d\left(A'_i v', A'_i w'\right) \leq \tau d\left(v', w'\right) \quad \text{for all } i \in \{1, \ldots, k\}, \tag{2.3}$$

$$c_0^{-1} \angle (v, w) \leq d\left(v', w'\right) \leq c_0 \angle (v, w) . \tag{2.4}$$

If M is forward-invariant for $\{A_1, \ldots, A_k\}$ then $M_c = \overline{(\mathbb{R}^2_* \setminus M)}$ is forward-invariant for $\{A_1^{-1}, \ldots, A_k^{-1}\}$. Moreover,

$$e'_1(\omega) = \bigcap_{n=1}^{\infty} A'_{\omega_{-1}} \cdot \ldots \cdot A'_{\omega_{-n}}(M')$$

and

$$e'_2(\omega) = \bigcap_{n=1}^{\infty} (A'_{\omega_{n-1}} \cdot \ldots \cdot A'_{\omega_0})^{-1}(M'_c).$$

Let $\omega = (\omega_-, \omega_+)$, where $\omega_- = \{\omega_i, i \leq -1\}$. Then $e_1(\omega) = e_1(\omega_-)$, $e_2(\omega) = e_2(\omega_+)$. We have $e_1(\Sigma) \subset M$, $e_2(\Sigma) \subset M_c$.

For 2×2 matrices $\lambda(\omega, v) = \lambda(\omega)$ for all $v \notin e_2(\omega)$. As $e_2(\omega) \subset M_c$ for all $\omega \in \Sigma$, we get

$$\lambda(\omega, v) = \lambda(\omega)$$

for all $\omega \in \Sigma$ and $v \in M$.

Given $\omega \in \Sigma$, consider the pair $(e_1(\omega), e_2(\omega))$. This behaves very nicely under action of the shift:

$$(e'_1(\sigma\omega), e'_2(\sigma\omega)) = (A'_{\omega_0} e'_1(\omega), A'_{\omega_0} e'_2(\omega)). \tag{2.5}$$

We say that the cocycle $\{A_1, \ldots, A_k\}$ satisfies the *forward non-overlapping condition* if we can choose a forward-invariant multicone M in such a way that $A_i(M) \cap A_j(M) = \emptyset$ for $i \neq j$. It satisfies the *backward non-overlapping condition* if we can choose a forward-invariant multicone M such that $A_i^{-1}(M_c) \cap A_j^{-1}(M_c) = \emptyset$

for $i \neq j$. If the cocycle satisfies both forward and backward non-overlapping condition (not necessarily for the same multicone), we say it satisfies the *non-overlapping condition* (NOC). The NOC is not only a geometric condition, it has a dynamical meaning as well: it is a necessary and sufficient condition for the map $\omega \to (e_1(\omega), e_2(\omega))$ to be a bijection.

3 Statement of Results

The paradigm of *ergodic optimization* (see [12]) says that for typical potentials the optimizing orbits (sets $\{\omega; \lambda(\omega) = \lambda^{\pm}\}$) should have low dynamical complexity. This is true in the commutative case, see [8, 13]. In the noncommutative situation it is probably not true in general, at least for the joint spectral subradius. However, in the open set of cocycles dominating and satisfying NOC, it is satisfied for all (not just typical) cocycles.

We will define *upper* and *lower Mather sets* K^+, K^- for a dominated cocycle $\{A_1, \dots, A_k\}$ as follows: K^+ (resp. K^-) is the union of supports of all σ-invariant measures μ on Σ such that $\lambda(\mu) = \lambda^+$ (resp. λ^-).

Theorem 3.1 *For a dominated cocycle, the Mather sets K^+, K^- are compact, nonempty, and invariant under σ. Moreover, every measure μ supported on K^+ (resp. K^-) satisfies $\lambda(\mu) = \lambda^+$ (resp. λ^-).*

Our main result is the following:

Theorem 3.2 *For a dominated cocycle satisfying NOC, the Mather sets K^+, K^- have zero topological entropy under σ.*

Both assumptions of Theorem 3.2 are necessary. For example, the cocycle

$$A_1 = \begin{pmatrix} 3 & 0 \\ 0 & 1/3 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 3 & 0 \\ 1 & 1/3 \end{pmatrix}$$

satisfies domination and the forward NOC, but still it does not satisfy the assertion of Theorem 3.2 (in this case, $K^+ = K^- = \Sigma$). For cocycles not satisfying domination the situation for joint spectral subradius is even worse: if we restrict our attention to cocycles $\{A_1, \dots, A_k\} \in SL(2, \mathbb{R})^k$, among cocycles not satisfying domination one can find an open and dense subset for which there exists an invariant positive topological entropy subset of Σ for which the norms $|A_n(\omega)|$ are uniformly bounded for all n (this corresponds to $\lambda^- = 0$). That is, the ergodic optimization fails.

The behaviour of the joint spectral radius is unknown in this case, but by a long-standing conjecture the ergodic optimization holds.

4 Barabanov Functions and Proof of Theorem 3.1

If the cocycle $\{A_1, \ldots, A_k\}$ is *irreducible* (has no nontrivial invariant subspace) then one can construct a *Barabanov norm*, that is a norm $|\cdot|_B$ on $\mathbb{R}^2$ such that for any $v \in \mathbb{R}^2$

$$\max_i |A_i v|_B = e^{\lambda^+} \cdot |v|_B. \tag{4.1}$$

In fact, such a norm can be defined in a much more general situation (i.e. for any irreducible compact subset of $GL(n, \mathbb{R})$), see [2, 14].

Unfortunately, in general there cannot exist a norm satisfying the analogue of (4.1) for the joint spectral subradius. However, we are able to construct a replacement (based on a similar idea in [6]). Given a dominated cocycle $\{A_1, \ldots, A_k\}$ with a forward-invariant multicone M, a pair of functions $p^+, p^- : M \to \mathbb{R}$ will be called *Barabanov functions* if they have the following properties:

- Extremality: for all $v \in M$,

$$\max_{i\in\{1,\ldots,k\}} p^+(A_i v) = p^+(v) + \lambda^+ , \tag{4.2}$$

$$\min_{i\in\{1,\ldots,k\}} p^-(A_i v) = p^-(v) + \lambda^- ; \tag{4.3}$$

- log-homogeneity: for all $v \in M$, and $t \in \mathbb{R}_*$,

$$p^\pm(tv) = p^\pm(v) + \log |t| ; \tag{4.4}$$

- Regularity: there exists $c_1 > 0$ such that for all $v_1, v_2 \in M$,

$$p^\pm(v_1) - p^\pm(v_2) \le c_1 \angle(v_1, v_2) + \log |v_1| - \log |v_2| . \tag{4.5}$$

Theorem 4.1 *For any dominated cocycle* $\{A_1, \ldots, A_k\}$ *there exist Barabanov functions* p^+, p^-.

Proof For each i let

$$h_i(v) = \log \frac{|A_i v|}{|v|}.$$

This function does not change under multiplying v by a scalar, hence it can be defined on $\mathbb{P}^1$. Let c_2 be the common Lipschitz constant of all h_is:

$$\left|h_i(v') - h_i(w')\right| \le c_2 \angle(v, w) \quad \text{for all } i \in \{1, \ldots, k\}, \text{ for all } v, w \in \mathbb{R}^2_*.$$

Let $c_3 = c_0c_2/(1-\tau)$. Let $\mathbb{K}$ be the space of c_3-Lipschitz functions (in d) from M' to $\mathbb{R}$ endowed with sup metric. For $f \in \mathbb{K}$ let

$$(T^+f)(v') = \max_{i\in\{1,\dots,k\}} \left[f\left(A_i'v'\right) + h_i(v')\right],$$
$$(T^-f)(v') = \min_{i\in\{1,\dots,k\}} \left[f\left(A_i'v'\right) + h_i(v')\right].$$

One can check that $T^+, T^- : \mathbb{K} \to \mathbb{K}$. We also have

$$T^\pm(f+c) = c + T^\pm f,$$

hence we can define T^+, T^- on the quotient $\hat{\mathbb{K}}$ of $\mathbb{K}$ by the subspace of constant functions. $\hat{\mathbb{K}}$ is convex and (by Arzela-Ascoli) compact, hence T^+ and T^- have fixed points in $\hat{\mathbb{K}}$ that we will denote by f_0^+, f_0^-. That is, there exist constants β^+, β^- such that

$$T^\pm f_0^\pm = f_0^\pm + \beta^\pm.$$

This immediately implies that the functions

$$p^\pm(v) = f_0^\pm(v') + \log|v|$$

satisfy all the required properties of Barabanov functions, with $\beta^\pm$ in place of $\lambda^\pm$. The only thing left is to check that $\beta^\pm$ cannot be different from $\lambda^\pm$.

Let us present this argument for β^+. For any vector $v \in M$ there exists a (not necessarily unique) $\omega_1^+ \in \{1,\dots,k\}$ such that $p^+(A_{\omega_1^+}v) = p^+(v) + \beta^+$. We can then find ω_2^+ such that $p^+(A_{\omega_2^+}A_{\omega_1^+}v) = p^+(v) + 2\beta^+$, and so on. Thus, β^+ is the maximal growth rate of p^+ for any vector $v \in M$. At the same time, by log-homogeneity of Barabanov functions, $p^+(v)$ can differ from $\log|v|$ by at most a constant. Hence, the growth rate of p^+ must be the same as the growth rate of $\log|\cdot|$, and we are done. □

The statement of Theorem 3.1 follows easily (once again, we will only construct K^+). Above we constructed for any vector $v \in M$ a set of infinite sequences $\Omega^+(v) \subset \{1,\dots,k\}^{\mathbb{N}}$ such that for every $\omega \in \Omega^+(v)$

$$p^+\left(A_n(\omega)v\right) = p^+(v) + n\lambda^+.$$

Consider the set $K_0^+ \subset \Sigma$ of the following form: $\omega = (\omega_-, \omega_+)$ belongs to K_0^+ if and only if $\omega_+ \in \Omega^+(e_1(\omega_-))$. Clearly, $\sigma K_0^+ \subset K_0^+$. We define

$$K^+ = \bigcap_{j=0}^{\infty} \sigma^j K_0^+.$$

This set is nonempty and compact, and has the following property: let $\omega \in K^+$ and $j \in \mathbb{Z}$. Let $\sigma^j\omega = (\omega_-^{(j)}, \omega_+^{(j)})$. Then

$$\omega_+^{(j)} \in \Omega^+(e_1(\omega_-^{(j)})).$$

It follows that every measure supported on K^+ has the maximal growth of p^+. Vice versa, every measure giving maximal growth of p^+ must for almost every past ω_- give full probability to futures from $\Omega^+(e_1(\omega_-))$, hence it must be supported on K^+. As the growth of p^+ must be the same as the growth of the length of any vector from M, this proves that the constructed set K^+ is the Mather set.

5 Proof of Theorem 3.2

The strategy of the proof is quite simple. We consider the space $\{(e_1(\omega), e_2(\omega)); \omega \in K^\pm\}$ with the dynamics given by (2.5). We will use Barabanov functions and geometric arguments to prove that this dynamical system has zero entropy (this result does not use NOC, only domination). We will then use NOC to transport the entropy result back to the full shift (Σ, σ).

Let us start with a simple lemma.

Lemma 5.1 *Let $\omega = (\omega_-, \omega_+) \in K^\pm$. Choose any $x \in e_1(\omega_-)$. If $y \in M$ is such that $x - y \in e_2(\omega_+)$ then*

$$p^+(x) \le p^+(y) \quad \text{if } \omega \in K^+,$$

$$p^-(x) \ge p^-(y) \quad \text{if } \omega \in K^-.$$

Proof Consider the case $\omega \in K^+$, the other is analogous. As $y - x \in e_2(\omega_+)$,

$$p^+(A_n(\omega)x) - p^+(A_n(\omega)y) \to 0.$$

At the same time,

$$p^+(A_n(\omega)x) - p^+(x) = n\lambda^+ \ge p^+(A_n(\omega)y) - p^+(y).$$

□

Given vectors x_1, y_1, x_2, $y_2 \in \mathbb{R}^2_*$, no three of them collinear, we define their *cross-ratio*

$$[x_1, y_1; x_2, y_2] := \frac{x_1 \times x_2}{x_1 \times y_2} \cdot \frac{y_1 \times y_2}{y_1 \times x_2} \in \mathbb{R} \cup \{\infty\},$$

where $\times$ denotes the cross-product in $\mathbb{R}^2$, i.e. the determinant. The cross-ratio depends only on the directions of the four vectors, hence we can define it on $(\mathbb{P}^1)^4$. See [7, Section 6].

Applying Lemma 5.1 twice, we get

Lemma 5.2 *Let* $\omega, \tau \in K^{\pm}$. *Then*

$$|[e_1(\omega_-), e_1(\tau_-); e_2(\omega_+), e_2(\tau_+)]| \geq 1 \quad \text{if } \omega, \tau \in K^+,$$
$$|[e_1(\omega_-), e_1(\tau_-); e_2(\omega_+), e_2(\tau_+)]| \leq 1 \quad \text{if } \omega, \tau \in K^-.$$

Proof We will consider the case $\omega, \tau \in K^+$, the other one is analogous. We choose $x_1 \in e_1(\omega_-), x_2 \in e_2(\omega_+), y_1 \in e_1(\tau_-), y_2 \in e_2(\tau_+)$. We can write

$$x_1 = \alpha x_2 + \beta y_1 \quad \text{and} \quad y_1 = \gamma y_2 + \delta x_1.$$

Applying Lemma 5.1 twice, we get

$$p^+(x_1) \leq p^+(\beta y_1) \leq p^+(\beta\delta x_1) = p^+(x_1) + \log|\beta\delta|.$$

Hence, $|\beta\delta| \geq 1$. Substituting

$$\beta = \frac{x_1 \times x_2}{y_1 \times x_2} \quad \text{and} \quad \delta = \frac{y_1 \times y_2}{x_1 \times y_2}$$

we obtain the assertion. □

We now use the hyperbolic geometry representation of $\mathbb{P}^1$. We identify the point $(\cos\theta, \sin\theta) \in \mathbb{P}^1$ with $e^{2\theta i}$ on the boundary $\partial\mathbb{D}$ of the unit disk $\mathbb{D}$. We endow $\mathbb{D}$ with the Poincaré hyperbolic metric. Given two points $x, y \in \partial\mathbb{D}$, we consider their connecting geodesic $\vec{xy} \in \mathbb{D}$.

There are three possible types of configurations of a 4-tuple of distinct points $(x_1, y_1; x_2, y_2)$ in $\mathbb{P}^1$ (see Fig. 1):

- *Antiparallel*: $x_1 < y_2 < y_1 < x_2 < x_1$ for some cyclic order $<$ on $\mathbb{P}^1$,
- *Coparallel*: $x_1 < y_1 < y_2 < x_2 < x_1$ for some cyclic order $<$ on $\mathbb{P}^1$,
- *Crossing*: $x_1 < y_1 < x_2 < y_2 < x_1$ for some cyclic order $<$ on $\mathbb{P}^1$.

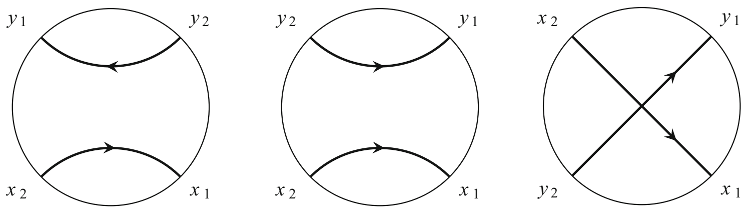


Fig. 1 From left to right: antiparallel, coparallel, and crossing configuration

Two geodesics $\overrightarrow{x_2x_1}$ and $\overrightarrow{y_2y_1}$ with distinct endpoints are called *antiparallel*, *coparallel*, or *crossing* according to the configuration of the 4-tuple $(x_1, y_1; x_2, y_2)$.

In terms of the cross-ratio, configurations can be expressed as follows.

Lemma 5.3 *The configuration of a* 4*-tuple* $(x_1, y_1; x_2, y_2)$ *in* $\mathbb{P}^1$ *is*

- Antiparallel *iff* $[x_1, y_1; x_2, y_2] < 0$,
- Coparallel *iff* $0 < [x_1, y_1; x_2, y_2] < 1$,
- Crossing *iff* $[x_1, y_1; x_2, y_2] > 1$.

Hence, Lemma 5.2 implies that for two sequences $\omega, \tau \in K^{\pm}$ the corresponding geodesics $e_1(\omega_-)\overrightarrow{\ }e_2(\omega_+), e_1(\tau_-)\overrightarrow{\ }e_2(\tau_+)$ cannot be in coparallel (if $\omega, \tau \in K^+$) or crossing (if $\omega, \tau \in K^-$) configuration.

We will not formulate the last part of the proof for the dynamical system acting on pairs $(e_1(\omega_-), e_2(\omega_+))$ but directly for $(K^{\pm}, \sigma)$. We recall that NOC guarantees that the two systems are conjugated. For K^+, K^- let us consider the sets of pasts with more than one future and sets of futures with more than one past. Formally, consider

$$N_1^+ = \{\omega_-;\ \text{there exists more than one } \omega_+ \text{ such that } (\omega_-, \omega_+) \in K^+\}, \tag{5.1}$$

$$N_2^+ = \{\omega_+;\ \text{there exists more than one } \omega_- \text{ such that } (\omega_-, \omega_+) \in K^+\}. \tag{5.2}$$

We define N_1^-, N_2^- analogously.

Lemma 5.4 *The sets* $N_1^+, N_2^+, N_1^-, N_2^-$ *are countable.*

Proof Consider N_1^+ first (the case of N_2^+ is analogous). Let $\omega_- \in N_1^+$. Denote by $I^+(\omega_-)$ the convex hull (taken in $\mathbb{P}^1 \setminus \{e_1(\omega_-)\}$) of the points $e_2(\omega_+)$ for ω_+ such that $(\omega_-, \omega_+) \in K^+$. Then for different $\omega_-, \tau_- \in N_1^+$ the intervals $I^+(\omega), I^+(\tau_-)$ have disjoint interiors. Indeed, otherwise some pairs of geodesics $e_1(\omega_-)\overrightarrow{\ }e_2(\omega_+), e_1(\tau_-)\overrightarrow{\ }e_2(\tau_+)$ would be in coparallel configuration, see Fig. 2.

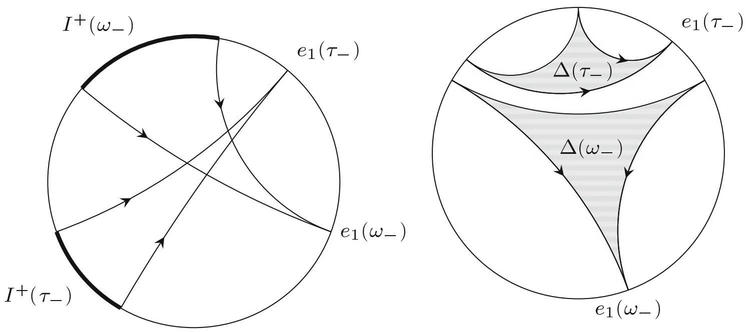


Fig. 2 Two cases in the proof of Lemma 5.4: disjoint arcs and disjoint geodesic triangles

Consider now the case N_1^- (or N_2^-). For any $\omega_- \in N_1^-$ we construct $I^-(\omega_-)$ analogously to $I^+(\omega_-)$ above, and then we construct the geodesic triangle $\Delta(\omega_-)$ with vertices $e_1(\omega_-)$ and the two endpoints of $I^-(\omega_-)$. Then for any two sequences $\omega_-, \tau_- \in N_1^-$ the triangles $\Delta(\omega_-), \Delta(\tau_-)$ have disjoint interiors (otherwise some pair of geodesics $\overrightarrow{e_1(\omega_-)e_2(\omega_+)}, \overrightarrow{e_1(\tau_-)e_2(\tau_+)}$ would have to be in crossing configuration), see Fig. 2.

The assertion follows by the separability of $\partial\mathbb{D}$ and $\mathbb{D}$. □

Thus, in either K^+ or K^- every past (except countably many) has a unique future and every future (except countably many) has a unique past. Such sets have zero topological entropy:

Lemma 5.5 *Let K be a compact σ-invariant subset of a two-sided shift. Define N_1, N_2 as in (5.1), (5.2). If N_1 and N_2 are countable then K has zero topological entropy.*

Proof It is enough to prove that every ergodic invariant measure has zero metric entropy. The atomic measures have entropy zero. The nonatomic measures do not see N_1, N_2, hence the past uniquely determines the future (and vice versa). This means that the conditional entropy of the generating partition with respect to the past/future is zero. □

6 Open Questions

There are many open questions. In particular:

- What happens for more general potentials (i.e. not piecewise constant)?
- What happens for more general base systems (for example, for subshifts of finite type)?
- What happens for matrices of size greater than 2×2?
- What happens in the generic situation?

References

1. A. Avila, J. Bochi, C. Yoccoz, Uniformly hyperbolic finite-valued $SL(2,\mathbb{R})$-cocycles. Comment. Math. Helv. **85**, 813–884 (2010)
2. N.E. Barabanov, On the Lyapunov exponent of discrete inclusions I. Autom. Remote Control **49**, 152–157 (1988)
3. L. Barreira, *Thermodynamic Formalism and Applications to Dimension Theory*. Progress in Mathematics, vol. 294 (Birkhäuser/Springer, Basel, 2011)
4. J. Bochi, N. Gourmelon, Some characterizations of domination. Math. Z. **263**, 221–231 (2009)
5. J. Bochi, M. Rams, The entropy of Lyapunov-optimizing measures of some metric cocycles (preprint)

6. T. Bousch, J. Mairesse, Asymptotic height optimization for topical IFS, Tetris heaps, and the finiteness conjecture. J. Am. Math. Soc. **15**, 77–111 (2002)
7. H. Busemann, P.J. Kelly, *Projective Geometry and Projective Metrics* (Academic, New York, 1953)
8. G. Contreras, Ground states are generically a periodic orbit (preprint)
9. D.-j. Feng, W. Huang, Lyapunov spectrum of asymptotically sub-additive potentials. Comm. Math. Phys. **297**, 1–43 (2010)
10. D.-J. Feng, P. Shmerkin, Non-conformal repellers and the continuity of pressure for matrix cocycles. Geom. Funct. Anal. **24**, 1101–1128 (2014)
11. R. Jungers, *The Joint Spectral Radius: Theory and Applications*. Lecture Notes in Control and Information Sciences, vol. 385 (Springer, Berlin, 2009)
12. O. Jenkinson, Ergodic optimization. Discr. Contin. Dyn. Syst. **15**, 197–224 (2006)
13. I.D. Morris, Maximizing measures of generic Hölder functions have zero entropy. Nonlinearity **21**, 993–1000 (2008)
14. F. Wirth, The generalized spectral radius and extremal norms. Linear Algebra Appl. **342**, 17–40 (2002)

Part III
Analysis and Algebra on Fractals

Poincaré Functional Equations, Harmonic Measures on Julia Sets, and Fractal Zeta Functions

Peter J. Grabner

Abstract We collect results from earlier work with G. Derfel and F. Vogl, which led to the proof of the existence of a meromorphic continuation of the fractal zeta function for certain self-similar fractals admitting spectral decimation. We explain the connection to classical functional equations occurring in the theory of polynomial iteration, namely Poincaré's and Böttcher's equations, as well as properties of the harmonic measure on the underlying Julia set. Furthermore, we comment on some more recent developments based on the work of N. Kajino and state a conjecture related to our approach via functional equations.

Keywords Fractal zeta function • Poincaré functional equation • Böttcher functional equation

1 Introduction

Connections between the analysis on fractals and the iteration of rational functions were discovered in the earliest publications on diffusion processes on certain self-similar sets, such as the Sierpiński gasket (see, for instance [3, 38]). The connection stems from the fact that time on the successive approximations of the fractal is modelled by a branching process. The relation of branching processes to the iteration of holomorphic functions is known for a long time (see [19]).

More precisely, in order to obtain a diffusion on a fractal, define a sequence of random walks on approximating graphs and synchronise time so that the limiting

The author is supported by the Austrian Science Fund (FWF): projects F5503 (part of the Special Research Program (SFB) "Quasi-Monte Carlo Methods: Theory and Applications"), W1230 (Doctoral Program "Discrete Mathematics"), and I1136 (French-Austrian international cooperation "Fractals and Numeration").

P.J. Grabner (✉)
Institut für Analysis und Computational Number Theory (Math A), NAWI-Graz, Technische Universität Graz, Steyrergasse 30, 8010 Graz, Austria
e-mail: peter.grabner@tugraz.at

C. Bandt et al. (eds.), *Fractal Geometry and Stochastics V*, Progress in Probability 70, DOI 10.1007/978-3-319-18660-3_10

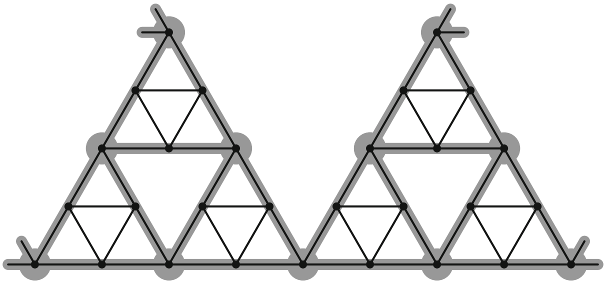


Fig. 1 The Sierpiński graph G in *black* with the graph G_F in *grey*. The corresponding set F consists of the *grey vertices*

process is non-constant and continuous. This was the first approach to the diffusion process on the Sierpiński gasket given in [3, 14, 26] and later generalised to other "nested fractals" in [29]. In our description we will follow the lines of definition of self-similar graphs given in [24, 25] and adapt it for our purposes.

We consider a graph $G = (V(G), E(G))$ with vertices $V(G)$ and undirected edges $E(G)$ denoted by $\{x, y\}$. We assume throughout that G does not contain multiple edges nor loops. For $C \subset V(G)$ we call ∂C the vertex boundary, which is given by the set of vertices in $V(G) \setminus C$, which are adjacent to a vertex in C. For $F \subset V(G)$ we define the reduced graph G_F by $V(G_F) = F$ and $\{x, y\} \in E(G_F)$, if x and y are in the boundary of the same component of $V(G) \setminus F$. This requires that removing the set F disconnects the graph G into different components.

The following definition is taken from [25]. It is motivated by the properties of the infinite Sierpiński gasket (see Fig. 1). Furthermore, it will turn out that this definition of self-similarity of a graph is reflected by according functional equations for the Green function (the generating function of the transition probabilities) and by rational function relations between the eigenvalues of the transition Laplace operator, which will be exploited later.

Definition 1.1 A connected infinite graph G is called self-similar with respect to $F \subset V(G)$ and $\varphi : V(G) \to V(G_F)$, if

1. No vertices in F are adjacent in G,
2. The intersection of the boundaries of two different components of $V(G) \setminus F$ does not contain more than one point,
3. φ is an isomorphism of G and G_F.

A random walk on G is given by transition probabilities $p(x, y)$, which are positive, if and only if $\{x, y\} \in E(G)$. For a trajectory $(Y_n)_{n\in\mathbb{N}_0}$ of this random walk with $Y_0 = x \in F$ we define stopping times recursively by

$$T_{m+1} = \min\{k > T_m \mid Y_k \in F \setminus \{Y_{T_m}\}\}, \quad T_0 = 0.$$

Then $(Y_{T_m})_{m\in\mathbb{N}_0}$ is a random walk on G_F. Since the underlying graphs G and G_F are isomorphic, it is natural to require that $(\varphi^{-1}(Y_{T_m}))_{m\in\mathbb{N}_0}$ is the same stochastic process as $(Y_n)_{n\in\mathbb{N}_0}$. This requires the validity of equations for the basic transition probabilities

$$\mathbb{P}\left(Y_{T_{n+1}} = \varphi(y) \mid Y_{T_n} = \varphi(x)\right) = \mathbb{P}\left(Y_{n+1} = y \mid Y_n = x\right) = p(x,y). \tag{1.1}$$

These are usually non-linear rational equations for the transition probabilities $p(x,y)$. The existence of solutions of these equations has been the subject of several investigations, and we refer to [31–33, 41] for further details.

The process $(Y_n)_{n\in\mathbb{N}_0}$ on G and its "shadow" $(Y_{T_n})_{n\in\mathbb{N}_0}$ on G_F are equal, but they are on a different time scale. Every transition $Y_{T_n} \to Y_{T_{n+1}}$ on G_F comes from a path $Y_{T_n} \to Y_{T_n+1} \cdots \to Y_{T_{n+1}-1} \to Y_{T_{n+1}}$ in a component of $V(G) \setminus F$. The *time scaling factor* between these processes is given by

$$\lambda = \mathbb{E}(T_{n+1} - T_n) = \mathbb{E}(T_1).$$

This factor is ≥ 2 by assumption (1) on F. More precisely, the relation between the transition time on G_F and the transition time on G is given by a super-critical ($\lambda > 1$) branching process, which replaces an edge $\{\varphi(x), \varphi(y)\} \in G_F$ by a path in G connecting the points x and y without visiting a point in $V(G) \setminus F$ (except for x, and for y in the last step).

In order to obtain a process on a fractal in $\mathbb{R}^d$, we assume further that G is embedded in $\mathbb{R}^d$ (i.e. $V(G) \subset \mathbb{R}^d$). The self-similarity of the graph is carried over to the embedding by assuming that there exists a $\beta > 1$ (the *space scaling factor*) such that $F = V(G_F) = \beta V(G)$. The fractal limiting structure is then given by

$$Z_G = \overline{\bigcup_{n=0}^{\infty} \beta^{-n} V(G)}.$$

Iterating this graph decimation we obtain a sequence $G_k = (\beta^{-k}V(G), E(G))$ of (isomorphic) graphs on different scales. The random walks $(Y_n^{(k)})_{n\in\mathbb{N}_0}$ on G_k (see Fig. 2) are connected by time scales with the scaling factor λ. From the theory of branching processes (cf. [19]) it follows that the time on level k scaled by λ^{-k} tends to a random variable. From this it follows that $\beta^{-k}Y_{\lfloor t\lambda^k \rfloor}$ weakly tends to a (continuous time) stochastic process $(X_t)_{t\geq 0}$ on the fractal Z_G. Notice, that β has to be chosen so that the limiting process $(X_t)_{t\geq 0}$ is continuous and not constant; thus there is of course only one (intrinsic) choice for β.

On the level of generating functions, the transition between the random walks on the graphs G_k and G_{k+1} is encoded by the relation

$$G(x,y \mid z) = g(z)G(\varphi(x), \varphi(y) \mid \psi(z)) \tag{1.2}$$

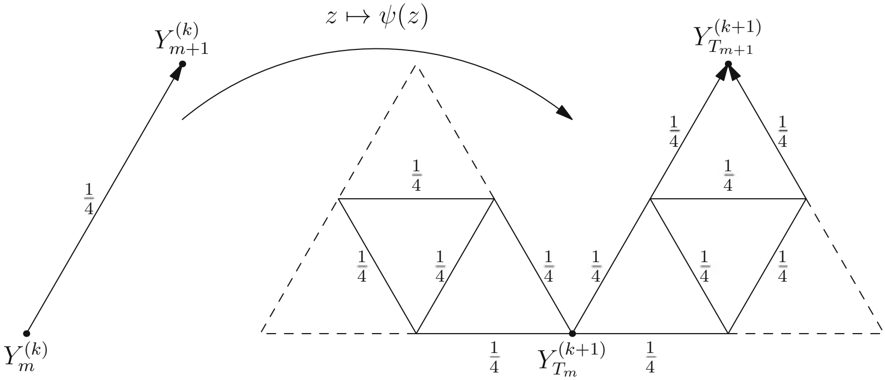


Fig. 2 Transition between $Y_m^{(k)}$ and $Y_{T_m}^{(k+1)}$

for the Green function

$$G(x, y \mid z) = \sum_{n=0}^{\infty} p_n(x, y) z^n,$$

where $p_n(x, y)$ denotes the n-step transition probability between x and y (cf. [15, 16, 25]). The generating function g encodes paths starting and ending in x without visiting any other point of F, whereas $\psi(z)$ is the probability generating function of all paths starting in a point of $a \in F$, ending in a point of $b \in F$, $b \neq a$, without visiting any point of F different from a, except for the last step.

The Laplace operator on Z_G is then defined as the infinitesimal generator of the semigroup of operators given by

$$A_t f(x) = \mathbb{E}_x f(X_t),$$

namely

$$\triangle f = \lim_{t \to 0+} \frac{A_t f - f}{t}, \tag{1.3}$$

defined for functions f, for which the limit exists.

It has been first observed by Fukushima and Shima [13, 42, 43] that the eigenvalues of the Laplacian on the Sierpiński gasket and its higher dimensional analogues exhibit the phenomenon of *spectral decimation* (see also earlier work by Bellissard [5, 6]). Later on, spectral decimation for more general fractals has been studied by Malozemov, Strichartz, and Teplyaev [1, 2, 30, 45, 47, 48].

Definition 1.2 (Spectral decimation) The Laplace operator on a p. c. f. self-similar fractal Z_G admits *spectral decimation*, if there exists a rational function R, a finite set A and a constant $\lambda > 1$ such that all eigenvalues of $\triangle$ can be written in the form

$$\lambda^m \lim_{n\to\infty} \lambda^n R^{(-n)}(\{w\}), \quad w \in A, \quad m \in \mathbb{N} \tag{1.4}$$

where the preimages of w under n-fold iteration of R have to be chosen such that the limit exists. Furthermore, the multiplicities $\beta_m(w)$ of the eigenvalues depend only on w and m, and the generating functions of these multiplicities are rational.

The fact that all eigenvalues of $\triangle$ are negative real implies that the Julia set of R has to be contained in the negative real axis. We will exploit this fact later.

The function R occurring in the definition of spectral decimation is conjugate to the function ψ occurring in (1.2) by a linear fractional transformation ξ, i.e. $R = \xi \circ \psi \circ \xi^{-1}$. In some cases such as the higher dimensional Sierpiński gaskets, the rational function R is a polynomial. This is the case that will be discussed further in this paper.

2 Polynomial Iteration

In order to discuss the consequences of spectral decimation further, we need to introduce some concepts and notation from the iteration theory of polynomials. Throughout, we will denote by $p^{(n)}$ the n-fold iterate of the (polynomial) function p, i.e.

$$p^0(z) = z, \quad p^{(n+1)}(z) = p\,(p^{(n)}(z)). \tag{2.1}$$

Let p be a real polynomial of degree d. We always assume that $p(0) = 0$ and $p'(0) = a_1 = \lambda$ with $|\lambda| > 1$. We refer to [4, 34] as general references for complex dynamics.

We denote the Riemann sphere by $\mathbb{C}_\infty$ and consider p as a map on $\mathbb{C}_\infty$. We recall that the Fatou set $\mathcal{F}(p)$ is the set of all $z \in \mathbb{C}_\infty$ which have an open neighbourhood U such that the sequence $(p^{(n)})_{n\in\mathbb{N}}$ is equicontinuous on U in the chordal metric on $\mathbb{C}_\infty$. By definition $\mathcal{F}(p)$ is open. We will especially need the component of ∞ of $\mathcal{F}(p)$ given by

$$\mathcal{F}_\infty(p) = \left\{z \in \mathbb{C} \mid \lim_{n\to\infty} p^{(n)}(z) = \infty\right\}, \tag{2.2}$$

as well as the basin of attraction of a finite attracting fixed point w_0 ($p(w_0) = w_0$, $|p'(w_0)| < 1$)

$$\mathcal{F}_{w_0}(p) = \left\{z \in \mathbb{C} \mid \lim_{n\to\infty} p^{(n)}(z) = w_0\right\}. \tag{2.3}$$

The complement of the Fatou set is the Julia set $\mathcal{J}(p) = \mathbb{C}_\infty \setminus \mathcal{F}(p)$.

The filled Julia set is given by

$$\mathcal{K}(p) = \left\{z \in \mathbb{C}_\infty \mid (p^{(n)}(z))_{n\in\mathbb{N}} \text{ is bounded}\right\} = \mathbb{C}_\infty \setminus \mathcal{F}_\infty(p). \tag{2.4}$$

Furthermore, it is known that (cf. [12])

$$\partial\mathcal{K}(p) = \partial\mathcal{F}_\infty(p) = \mathcal{J}(p). \tag{2.5}$$

This relation only holds for polynomials; for the iteration of general rational functions the situation is much more complicated.

3 Poincaré's Functional Equation

We now want to analyse Eq. (1.4) further, assuming that $R = p$, a polynomial of degree d with a fixed point at 0 with $p'(0) = \lambda > 1$. Let z be a complex number obtained as a limit

$$\lim_{n\to\infty} \lambda^n p^{(-n)}(\{w\}); \tag{3.1}$$

this means that

$$\lim_{n\to\infty} p^{(n)}(\lambda^{-n}z) = w. \tag{3.2}$$

It is a well known fact from the iteration theory of polynomials that the function sequence $(p^{(n)}(\lambda^{-n}z))_n$ converges uniformly on compact sets to an entire function $\Phi(z)$. This function satisfies the Poincaré functional equation

$$\Phi(\lambda z) = p(\Phi(z)), \quad \Phi(0) = 0, \quad \Phi'(0) = 1. \tag{3.3}$$

The function Φ provides a linearisation of the action of p around 0 and was studied intensively since the fundamental work of H. Poincaré [36, 37]. The order of this function and precise asymptotic information about its maximal function

$$M_\Phi(r) = \max_{|z|=r} |\Phi(z)| \tag{3.4}$$

were derived in [49, 50]. In [8, 9] a complete asymptotic expansion valid in certain angular regions of the complex plane could be obtained. This was used in [10] to give an analytic continuation of the spectral ζ-function

$$\zeta_\triangle(s) = \sum_{-\triangle u = \mu u} \mu^{-s} \tag{3.5}$$

of the Laplace operator to the whole complex plane. For future reference, we denote the abscissa of convergence of this Dirichlet series by $\frac{1}{2}d_S$, the *spectral dimension*. The factor $\frac{1}{2}$ is added by convention so that the classical result of H. Weyl [51] for the asymptotic expansion of the eigenvalue counting function on a compact d-dimensional manifold Ω

$$N_{\triangle}(x) = \sum_{\substack{-\triangle u = \mu u \\ \mu < x}} 1, \tag{3.6}$$

namely (ω_d is the volume of the d-dimensional unit ball)

$$N_{\triangle}(x) \sim \frac{\omega_d}{(2\pi)^d} \mathrm{vol}(\Omega) x^{\frac{d}{2}}$$

is reproduced as a special case.

The values z that can be obtained by (3.1) are exactly the solutions of the equation $\Phi(z) = w$. As is well known from the theory of entire functions (see [7]), the behaviour of the counting function of the number of solutions of $\Phi(z) = w$ in a circle of radius r is directly connected to the growth order of Φ, or more precisely, the maximal function $M_\Phi(r)$ in (3.4).

4 Böttcher's Functional Equation

As was pointed out in Sect. 3, the Poincaré-function $\Phi(z)$ given by (3.3) provides a local linearisation of the polynomial function p around its fixed point $z = 0$. The construction of this function as the limit (3.1) depends heavily on the fact that $|\lambda| > 1$ (*repelling fixed point*), where $\lambda = p'(0)$. A similar linearisation can be found for $0 < |\lambda| < 1$ (*attracting fixed point*); the case of an *indifferent fixed point* ($|\lambda| = 1$) is much more delicate and the existence of a local linearisation depends heavily on Diophantine conditions on the argument of λ (see [4, 34]). The case of vanishing derivative $\lambda = 0$ (*hyper-attracting fixed point*) leads to a different kind of linearisation, which shall be the subject of this section. Notice, that $z = \infty$ is such a fixed point for a polynomial of degree $d \geq 2$, if considered as a function on the Riemann sphere.

The Böttcher functional equation associated to the hyper-attracting fixed point ∞ of a polynomial $p(z) = a_d z^d + \cdots + a_0$ of degree $d \geq 2$ is given by

$$a_d (g(z))^d = g(p(z)). \tag{4.1}$$

The solution of this equation exists in some neighbourhood of ∞ and can be expressed as a Laurent series around ∞

$$g(z) = z + \sum_{n=0}^{\infty} \frac{c_n}{z^n}.$$

Furthermore, the sequence of functions $(a_d^{-\frac{1}{d-1}}(p^{(n)}(z))^{d^{-n}})_n$ converges uniformly to g on compact subsets of $\mathbb{C}_\infty$ contained in the domain of g (if the branches of the d-th roots are chosen accordingly).

The Böttcher function $g(z)$ admits the integral representation, which also provides an analytic continuation of g to any simply connected subset of $\mathbb{C} \setminus \mathcal{K}(p)$

$$g(z) = \exp\left(\int_{\mathcal{J}(p)} \log(z-x)\,\mathrm{d}\mu(x)\right), \tag{4.2}$$

where μ denotes the harmonic measure on $\mathcal{J}(p)$. The measure μ is the unique probability measure supported on $\mathcal{J}(p)$ minimising the logarithmic energy

$$\mathcal{E}(\nu) = \int_{\mathcal{J}(p)}\int_{\mathcal{J}(p)} \log\frac{1}{|z-w|}\,\mathrm{d}\nu(z)\,\mathrm{d}\nu(w)$$

(see [39]). For the measure μ the corresponding potential

$$U_\mu(z) = \int_{\mathcal{J}(p)} \log\frac{1}{|z-w|}\,\mathrm{d}\mu(w)$$

is constant on $\mathcal{K}(p)$; this constant equals $\frac{1}{d-1}\log a_d$, the logarithm of the capacity of $\mathcal{J}(p)$. This is also the value of the energy $\mathcal{E}(\mu)$.

The measure μ can be obtained as the weak limit of the sequence of measures

$$\frac{1}{d^n}\sum_{p^{(n)}(\xi)=x} \delta_\xi, \tag{4.3}$$

where x is an arbitrarily chosen point and δ_ξ denotes a unit point mass at ξ. The fact that (4.2) and (4.1) yield the same function, follows immediately from $p^*(\mu) = d\cdot\mu$.

Equation (4.2) can be used to obtain an analytic continuation of $g(z)$ to any simply connected subset of $\mathbb{C}_\infty \setminus \mathcal{K}(p)$. Furthermore, if $\mathcal{K}(p)$ is connected, $1/(a_d^{1/(d-1)}g(z))$ is the Riemann mapping, mapping $\mathbb{C}_\infty \setminus \mathcal{K}(p)$ to the unit circle.

The function $\log|g(z)|$ is the Green function for the logarithmic potential on $\mathcal{F}_\infty(p)$ and

$$\lim_{\substack{z\to z_0\\ z\in\mathcal{F}_\infty(p)}} |g(z)| = a_d^{-\frac{1}{d-1}} \Leftrightarrow z_0 \in \mathcal{J}(p).$$

In the case that $\mathcal{K}(p)$ is not connected, the mapping $g : \mathbb{C} \setminus \mathcal{K}(p) \to \mathbb{C}$ is much more complicated. For further details we refer to [28].

5 Asymptotic Behaviour of Poincaré Functions

Combining the solutions of the functional equations (3.3) and (4.1), we are now in the position to obtain an asymptotic expansion of the Poincaré function Φ for real values of λ inside angular regions, where Φ tends to ∞.

Consider the function $h(z) = g(\Phi(z))$ in an angular region

$$W_{\alpha,\beta} = \{z \in \mathbb{C} \setminus \{0\} \mid \alpha < \arg(z) < \beta\},$$

where Φ tends to ∞. Then h satisfies the functional equation

$$a_d h(z)^d = h(\lambda z),$$

which has the solution

$$h(z) = a_d^{-\frac{1}{d-1}} \exp\left(z^\rho F(\log_\lambda z)\right), \tag{5.1}$$

where $\rho = \log_\lambda d$, and F is a periodic function of period 1, which is holomorphic on the strip

$$\{z \in \mathbb{C} \mid \frac{\alpha}{\log\lambda} < \Im(z) < \frac{\beta}{\log\lambda}\}.$$

Furthermore, the fact that Φ tends to ∞ in $W_{\alpha,\beta}$ yields

$$\forall z \in W_{\alpha,\beta} : \Re(z^\rho F(\log_\lambda z)) > 0.$$

Writing

$$g^{(-1)}(w) = w + \sum_{n=0}^{\infty} \frac{b_n}{w^n},$$

we obtain the full asymptotic expansion

$$\Phi(z) = a_d^{-\frac{1}{d-1}} \exp\left(z^\rho F(\log_\lambda z)\right) + \sum_{n=0}^{\infty} b_n a_d^{\frac{n}{d-1}} \exp\left(-nz^\rho F(\log_\lambda z)\right) \tag{5.2}$$

valid for $z \in W_{\alpha,\beta}$. This derivation is the content of [9, Theorem 2.1].

Taking the logarithm of (5.1) and using the fact that $\Phi(z) = z + O(z^2)$ for $z \to 0$, we obtain

$$\log g(z) = \int_{\mathcal{J}(p)} \log(z-x)\,\mathrm{d}\mu(x) \sim -\frac{1}{d-1}\log a_d + z^\rho F(\log_\lambda z) + O(z^{2\rho}) \tag{5.3}$$

for $z \to 0$ in $W_{\alpha,\beta}$. On the other hand, taking the logarithm of (5.2), we get

$$\log \Phi(z) = -\frac{1}{d-1}\log a_d + z^\rho F(\log_\lambda z) + O(\exp(-z^\rho F(\log_\lambda z))) \tag{5.4}$$

for $z \to \infty$ again in $W_{\alpha,\beta}$. This means that the same periodic function F can be observed in the asymptotic behaviour of $\log g$ for $z \to 0$ and $\log \Phi$ for $z \to \infty$. The function F encodes properties of the Julia set $\mathcal{J}(p)$ in the following sense.

Theorem 5.1 ([9, Theorem 2.2]) *The periodic function F is constant, if and only if the polynomial is either linearly conjugate to z^d or to the Chebyshev polynomial of the first kind $T_d(z)$. In the first case the Julia set $\mathcal{J}(p)$ is a circle, in the second case the Julia set $\mathcal{J}(p)$ is a closed interval.*

Remark 1 It is known from [17] that the circle and the interval are the only cases of smooth Julia sets; these occur precisely for the polynomials described in the Theorem.

6 Fractal Zeta Functions

We now return to the study of the spectrum of the Laplacian $\triangle$ on a fractal admitting spectral decimation with the polynomial p in the sense of Definition 1.2. In this case the Julia set $\mathcal{J}(p)$ is contained in the negative real axis, which implies that $\rho = \log_\lambda d \leq \frac{1}{2}$ by [9, Theorem 4.1]. The Poincaré function Φ is thus an entire function of order $\rho \leq \frac{1}{2}$. Here, the case $\rho = \frac{1}{2}$ can only occur, if $\mathcal{J}(p)$ is an interval, or equivalently p is a Chebyshev polynomial. In the context of fractals with spectral decimation, this occurs, if the fractal is a compact interval viewed as a self-similar fractal. If this case is ruled out, we have $\rho < \frac{1}{2}$. Functions of order $< \frac{1}{2}$ are unbounded on every ray (see [7]). Furthermore, this together with the fact that Φ attains values in $\mathcal{J}(p) = \mathcal{K}(p)$ only for negative real arguments yields that

$$\lim_{\substack{z\to\infty \\ z\in W_{-\pi,\pi}}} \Phi(z) = \infty. \tag{6.1}$$

Especially, this implies that $\lim_{x\to+\infty} \Phi(x) = \infty$ and thus (5.2) holds for $z \to +\infty$ along the positive real axis.

Let $-\xi_\ell(w)$ ($\ell = 1, 2 \ldots$) denote the solutions of $\Phi(z) = w$; for $w = 0$, we set $\xi_0(0) = 0$ and $\xi_\ell(0) \neq 0$ for $\ell = 1, 2, \ldots$. Define

$$\Phi_0(z) = \frac{1}{z}\Phi(z) \quad \text{and} \quad \Phi_w(z) = 1 - \frac{1}{w}\Phi(z).$$

Then we have the following Hadamard product expansion

$$\Phi_w(z) = \prod_{\ell=1}^{\infty} \left(1 + \frac{z}{\xi_\ell(w)}\right). \tag{6.2}$$

Taking the Mellin transform of the logarithm of (6.2) yields

$$M_w(s) = \int_0^\infty \log(\Phi_w(x))x^{s-1}\,\mathrm{d}x = \frac{\pi}{s\sin\pi s}\sum_{\ell=1}^{\infty} \xi_\ell(w)^s \tag{6.3}$$

for $-1 < \Re(s) < -\rho$. The left inequality comes from the fact that $\log(\Phi_w(x)) = O(x)$ for $x \to 0$, whereas the right inequality comes from the behaviour of Φ for $x \to \infty$ given in (5.2): $\log(\Phi_w(x)) = O(x^\rho)$.

The functions

$$\zeta_{\Phi,w}(s) = \sum_{\ell=1}^{\infty} \xi_\ell(w)^{-s} \tag{6.4}$$

will be used to derive an expression for $\zeta_\triangle$ later. In order to obtain an analytic continuation of $\zeta_\triangle$ to the whole complex plane, we will need analytic continuations of the functions $\zeta_{\Phi,w}$. We will follow the lines of [11]; similar, but slightly different ideas were used in [10].

We consider the function

$$\Psi_w(z) = \frac{p(\Phi(z)) - w}{a_d(\Phi(z) - w)} = \frac{\Phi_w(\lambda z)}{a_d(-w)^{d-1}\Phi_w(z)^d}$$

for $w \neq 0$. Taking the logarithm, we obtain

$$\log\Psi_w(z) = \log\Phi_w(\lambda z) - d\log\Phi_w(z) - \log a_d - (d-1)\log(-w);$$

this function tends to 0 like $\exp(-cz^\rho)$ for $z \to +\infty$. Taking the Mellin transform and using standard methods to obtain analytic continuations of such transforms, we obtain (we indicate the region of validity of the equation in every line)

$$\begin{aligned}
&(\lambda^{-s} - d)M_w(s)\\
&= \int_0^\infty (\log \Phi_w(\lambda x) - d \log \Phi_w(x))\, x^{s-1}\, \mathrm{d}x \quad (\text{for} -1 < \Re s < -\rho)\\
&= \int_0^1 (\log \Phi_w(\lambda x) - d \log \Phi_w(x))\, x^{s-1}\, \mathrm{d}x - (\log a_d + (d-1)\log(-w))\frac{1}{s}\\
&\quad + \int_1^\infty (\log \Phi_w(\lambda x) - d \log \Phi_w(x) - \log a_d - (d-1)\log(-w))\, x^{s-1}\, \mathrm{d}x \quad (\text{for } \Re s > -1)\\
&= \int_0^\infty \log(\Psi_w(x)) x^{s-1}\, \mathrm{d}x \quad (\text{for } \Re s > 0).
\end{aligned}$$

The above computation shows that $M_w(s)$ has a simple pole at $s = 0$ with residue

$$\operatorname*{Res}_{s=0} M_w(s) = \frac{\log a_d}{d-1} + \log(-w).$$

Furthermore, it provides an analytic continuation of $M_w(s)$ to the half-plane $\Re s > 0$; the second line also gives the analytic continuation to the half-plane $\Re s > -1$. Using (6.3) gives an analytic continuation of $\zeta_{\Phi,w}(s)$ to the half-plane $\Re s < 0$

$$\zeta_{\Phi,w}(s) = \frac{s \sin \pi s}{\pi(\lambda^s - d)} \int_0^\infty \log(\Psi_w(x)) x^{-s-1}\, \mathrm{d}x.$$

From this we derive the existence of "trivial zeros" $\zeta_{\Phi,w}(-m) = 0$ (for $m \in \mathbb{N}_0$). Notice, that the simple pole of $M_w(s)$ at $s = 0$ is cancelled by the double zero of $s \sin \pi s$. Observing this, we also obtain

$$\zeta'_{\Phi,w}(0) = -\frac{\log a_d}{d-1} - \log(-w).$$

Similar computations yield the analytic continuation of $\zeta_{\Phi,0}$ to the whole complex plane; this function has "trivial" zeros $\zeta_{\Phi,0}(-m) = 0$ (for $m \in \mathbb{N}$) and

$$\zeta_{\Phi,0}(0) = 1, \quad \zeta'_{\Phi,0}(0) = -\frac{\log a_d}{d-1}.$$

Simple poles of $\zeta_{\Phi,w}(s)$ can occur only at the solutions of $\lambda^s = d$, namely $s = \rho + 2k\pi i/\log\lambda$ ($k \in \mathbb{Z}$). These poles are in correspondence with the growth order of Φ, which implies that there is a pole at $s = \rho$. The other poles for $k \neq 0$ only occur, if the periodic function F in (5.2) is not constant. Theorem 5.1 characterises the polynomials, for which the periodic function F is constant.

We now use the assumption that the multiplicities $\beta_m(w)$ of the eigenvalues $\lambda^m\xi_\ell(w)$ have a rational generating function (see Definition 1.2). Let

$$B_w(x) = \sum_{m=0}^{\infty} \beta_m(w)x^m.$$

Then using our knowledge on the eigenvalues of $\triangle$ together with our assumptions from the definition of spectral decimation, we obtain

$$\zeta_\triangle(s) = \sum_{w\in A} B_w(\lambda^{-s})\zeta_{\Phi,w}(s). \tag{6.5}$$

This expression provides the analytic continuation of the spectral zeta function to the whole complex plane.

If $\rho < \frac{1}{2}d_S$ then all the functions $\zeta_{\Phi,w}(s)$ are holomorphic in a half-plane $\Re s > \frac{1}{2}d_S - \varepsilon$ for some $\varepsilon > 0$. On the other hand, $\zeta_\triangle(s)$ has a simple pole at $s = \frac{1}{2}d_S$ by the fact that $N_\triangle(x) \asymp x^{\frac{1}{2}d_S}$ (see [23]). Thus at least one of the rational functions $B_w(x)$ has to have a pole at $x = \lambda^{-\frac{1}{2}d_S}$. Since all the rational functions B_w have positive power series coefficients (the multiplicities of the eigenvalues), there can be no cancellation of poles, which implies that the functions B_w can have at most a simple pole at $x = \lambda^{-\frac{1}{2}d_S}$. Let W denote the set of all $w \in A$, for which the corresponding function B_w has a (simple) pole at $x = \lambda^{-\frac{1}{2}d_S}$. Then we write the Laurent expansion of $B_w(x)$ around $x = \lambda^{-\frac{1}{2}d_S}$ in the form

$$B_w(x) = \frac{c_1(w)}{1 - x\lambda^{\frac{1}{2}d_S}} + \cdots .$$

This implies that $c_1(w) > 0$ by the combinatorial interpretation of B_w. Then the Dirichlet series

$$\eta(s) = \sum_{w\in W} c_1(w)\zeta_{\Phi,w}(s)$$

has positive coefficients. By [27, Theorem 9.5,p. 184] this implies that $\eta(\frac{1}{2}d_S + ik\tau) = 0$ cannot hold for fixed $\tau > 0$ and all $k \in \mathbb{Z}\setminus\{0\}$. Thus the function

$$\sum_{w\in W} B_w(\lambda^{-s})\zeta_{\Phi,w}(s)$$

has a simple pole at $s = \frac{1}{2}d_S$ and at least two non-real poles on the line $\Re s = \frac{1}{2}d_S$. The remaining summands in (6.5) do not have poles on the line $\Re s = \frac{1}{2}d_S$; thus the function $\zeta_\triangle(s)$ has at least two non-real poles on this line.

As a conclusion, we have reached the following theorem (see [10, Theorem 9]).

Theorem 6.1 *Let Z_G be a p. c. f. self-similar compact fractal, whose Laplace operator $\triangle$ admits spectral decimation in the sense of Definition 1.2 with a polynomial of degree d. Then the Dirichlet generating function of the eigenvalues of $\triangle$*

$$\zeta_\triangle(s) = \sum_{-\triangle u=\mu u} \frac{1}{\mu^s},$$

has a meromorphic continuation to the whole complex plane with poles contained in a finite union of sets $\{\rho_k + 2\pi i m\sigma \mid m \in \mathbb{Z}\}$, where $\sigma = \frac{1}{\log\lambda}$ and λ is the parameter coming from spectral decimation. There is a simple pole at $s = \frac{1}{2}d_S$. If $\log_\lambda d < \frac{1}{2}d_S$ then $\zeta_\triangle(s)$ has at least two non-real poles on the line $\Re s = \frac{1}{2}d_S$.

Remark 2 The case of $G = [0,1]$ which gives the Riemann zeta function and has $\log_\lambda d = \frac{1}{2}d_S$ shows that the condition $\log_\lambda d < \frac{1}{2}d_S$ is needed for the last assertion. The case $\log_\lambda d > \frac{1}{2}d_S$ cannot occur.

7 Consequences and a Conjecture

We introduce one further notion in connection with the diffusion on a fractal, namely the trace of the heat operator

$$P(t) = \mathrm{Tr}(A_t) = \mathrm{Tr}(e^{t\triangle}) = \sum_{-\triangle u=\mu u} e^{-\mu t}. \tag{7.1}$$

In the classical case of a Riemannian manifold studied by H. Weyl [51], the behaviour of this function for $t \to 0+$ was used to prove asymptotic relations for the eigenvalue counting function $N_\triangle$. Furthermore, precise information on the asymptotic behaviour of $P(t)$ for $t \to 0+$ can be used to prove that the spectral zeta function of $\triangle$ on a Riemannian manifold has an analytic continuation to the whole complex plane (see [35, 40]). In the case of a fractal with spectral decimation, we proceed in the opposite direction; starting from precise information on the eigenvalues we derive the existence of an analytic continuation of $\zeta_\triangle$ to the whole complex plane with the location of all poles, from which we conclude asymptotic information about $N_\triangle$ and $P(t)$. We sum this up by citing the following theorem.

Theorem 7.1 ([10, Theorem 10]) *Let Z_G be a p. c. f. self-similar compact fractal, whose Laplace operator $\triangle$ admits spectral decimation in the sense of Definition 1.2. Then the following are equivalent:*

1. *$\zeta_\triangle(s)$ has at least two non-real poles in the set* $\frac{1}{2}d_S + \frac{2\pi i}{\log\lambda}\mathbb{Z}$,
2. *The limit* $\lim_{x\to\infty} x^{-\frac{1}{2}d_S} N_\triangle(x)$ *does not exist, where* $N_\triangle(x)$ *denotes the eigenvalue counting function* (3.6),
3. *The limit* $\lim_{t\to 0+} P(t)t^{\frac{1}{2}d_S}$ *does not exist, where* $P(t)$ *denotes the trace of the heat kernel* (7.1).

Remark 3 Recently, N. Kajino [20–22] could prove an asymptotic expansion of the trace of the heat kernel on a p. c. f. fractal and also on the generalised Sierpiński carpet

$$P(t) = \sum_{k=0}^{n} t^{-\alpha_k} G_k(\log t) + O\left(\exp\left(-ct^{-\gamma}\right)\right) \quad \text{for} t \to 0+$$

for certain exponents $\alpha_0 > \alpha_1 > \cdots > \alpha_n \geq 0$, periodic continuous functions G_k ($k = 0, \ldots, n$), and $c, \gamma > 0$. This result was obtained without precise knowledge of the eigenvalues and properties of the zeta function. This was used in [44] to obtain an analytic continuation of the zeta function $\zeta_\triangle$ to the whole complex plane in these cases.

Remark 4 Theorems 6.1 and 7.1 together show that the limit $\lim_{x\to\infty} x^{-\frac{1}{2}d_S} N_\triangle(x)$ does not exist for fractals admitting spectral decimation with a polynomial of degree d and $\log_\lambda d < \frac{1}{2}d_S$.

More precisely, in the case that $\log_\lambda d < \frac{1}{2}d_S$ we obtain

$$N_\triangle(x) = x^{\frac{1}{2}d_S} Q(\log_\lambda x) + o\left(x^{\frac{1}{2}d_S}\right) \quad \text{for} x \to \infty$$

and

$$P(t) = t^{-\frac{1}{2}d_S} R(\log_\lambda t) + O\left(t^{-\frac{1}{2}d_S+\varepsilon}\right) \quad \text{for} t \to 0+$$

for some $\varepsilon > 0$ and for continuous periodic functions with period 1, Q and R (see [10, 11]).

Remark 5 In [18] it was shown that there exist gaps in the spectrum of the Laplacian if and only if the Julia set of the spectral decimation function R is totally disconnected. Spectral gaps (in the sense that there exists a subsequence, along which the quotient of consecutive eigenvalues stays bounded away from 1) yields uniform convergence of the Fourier series of continuous functions along the subsequence mentioned above (see [46]).

In the context of fractals the polynomials occurring for spectral decimation have a negative real Julia set $\mathcal{J}(p)$ (which is a Cantor set, except for the case when $\mathcal{J}(p)$ is an interval; this last case only occurs, if the underlying fractal itself is an interval). Nevertheless, the Poincaré and Böttcher functions can be defined and studied for any polynomial p of degree $d \geq 2$. This was done in [9]. There the asymptotic behaviour of the zero counting function of Φ

$$N_\Phi(x) = \sum_{\substack{|\xi|<x \\ \Phi(\xi)=0}} 1 \tag{7.2}$$

could be related to the behaviour of the harmonic measure of small balls around the origin, namely

Theorem 7.2 ([9, Theorem 5.2]) *Let Φ be the entire solution of* (3.3)*, and let $\rho = \log_\lambda d$. Then the limit*

$$\lim_{x\to\infty} x^{-\rho} N_\Phi(x) \tag{7.3}$$

exists, if and only if the limit

$$\lim_{t\to 0} t^{-\rho} \mu(B(0,t)) \tag{7.4}$$

exists.

We repeat the following conjecture about the existence of the limits (7.3) and (7.4)

Conjecture 7.3 ([9]) The limits (7.3) and (7.4) exist, if and only if the polynomial p is either linearly conjugate to a pure power or a Chebyshev polynomial of the first kind.

References

1. N. Bajorin, T. Chen, A. Dagan, C. Emmons, M. Hussein, M. Khalil, P. Mody, B. Steinhurst, A. Teplyaev, Vibration modes of $3n$-gaskets and other fractals. J. Phys. A **41**(1), 015101, (21pp) (2008)
2. N. Bajorin, T. Chen, A. Dagan, C. Emmons, M. Hussein, M. Khalil, P. Mody, B. Steinhurst, A. Teplyaev, Vibration spectra of finitely ramified, symmetric fractals. Fractals **16**, 243–258 (2008)
3. M.T. Barlow, E.A. Perkins, Brownian motion on the Sierpiński gasket. Probab. Theory Relat. Fields **79**, 543–623 (1988)
4. A.F. Beardon, *Iteration of Rational Functions*. Graduate Texts in Mathematics, vol. 132 (Springer, New York, 1991)
5. J. Bellissard, Stability and instability in quantum mechanics, in *Trends and Developments in the Eighties (Bielefeld, 1982/1983)*, ed.by S. Albeverio, Ph. Blanchard (World Science, Singapore, 1985), pp. 1–106
6. J. Bellissard, Renormalization group analysis and quasicrystals, in *Ideas and Methods in Quantum and Statistical Physics (Oslo, 1988)*, ed. by S. Albeverio, J.E. Fenstad, H. Holden, T. Lindstrøm (Cambridge University Press, Cambridge, 1992), pp. 118–148
7. R.P. Boas Jr., *Entire Functions* (Academic, New York, 1954)

8. G. Derfel, P.J. Grabner, F. Vogl, Asymptotics of the Poincaré functions, in *Probability and Mathematical Physics: A Volume in Honor of Stanislav Molchanov (Montreal)*, ed. by D. Dawson, V. Jaksic, B. Vainberg. CRM Proceedings and Lecture Notes, vol. 42 (Centre de Recherches Mathématiques, Providence 2007), pp. 113–130
9. G. Derfel, P.J. Grabner, F. Vogl, Complex asymptotics of Poincaré functions and properties of Julia sets. Math. Proc. Camb. Philos. Soc. **145**, 699–718 (2008)
10. G. Derfel, P.J. Grabner, F. Vogl, The Zeta function of the Laplacian on certain fractals. Trans. Am. Math. Soc. **360**, 881–897 (2008)
11. G. Derfel, P.J. Grabner, F. Vogl, Laplace operators on fractals and related functional equations. J. Phys. A **45**(46), 463001 (34pp) (2012)
12. K.J. Falconer, *Fractal Geometry: Mathematical Foundations and Applications* (Wiley, Hoboken, 2003)
13. M. Fukushima, T. Shima, On the spectral analysis for the Sierpiński gasket. J. Potential Anal. **1**, 1–35 (1992)
14. S. Goldstein, Random walks and diffusions on fractals, in *Percolation Theory and Ergodic Theory of Infinite Particle Systems (Minneapolis, 1984–1985)*, ed. by H. Kesten. IMA Volumes in Mathematics and Its Applications, vol. 8 (Springer, New York, 1987), pp. 121–129
15. P.J. Grabner, Functional iterations and stopping times for Brownian motion on the Sierpiński gasket. Mathematika **44**, 374–400 (1997)
16. P.J. Grabner, W. Woess, Functional iterations and periodic oscillations for the simple random walk on the Sierpiński graph. Stoch. Process. Appl. **69**, 127–138 (1997)
17. D.H. Hamilton, Length of Julia curves. Pac. J. Math. **169**, 75–93 (1995)
18. K.E. Hare, B.A. Steinhurst, A. Teplyaev, D. Zhou, Disconnected Julia sets and gaps in the spectrum of Laplacians on symmetric finitely ramified fractals. Math. Res. Lett. **19**(3), 537–553 (2012)
19. T.E. Harris, *The Theory of Branching Processes* (Springer, Berlin/New York, 1963)
20. N. Kajino, Spectral asymptotics for Laplacians on self-similar sets. J. Funct. Anal. **258**(4), 1310–1360 (2010)
21. N. Kajino, On-diagonal oscillation of the heat kernels on post-critically finite self-similar fractals. Probab. Theory Relat. Fields **156**(1–2), 51–74 (2013)
22. N. Kajino, Log-periodic asymptotic expansion of the spectral partition function for self-similar sets. Comm. Math. Phys. **328**(3), 1341–1370 (2014)
23. J. Kigami, M.L. Lapidus, Weyl's problem for the spectral distribution of Laplacians on p.c.f. self-similar fractals. Comm. Math. Phys. **158**, 93–125 (1993)
24. B. Krön, Green functions on self-similar graphs and bounds for the spectrum of the Laplacian. Ann. Inst. Fourier (Grenoble) **52**, 1875–1900 (2002)
25. B. Krön, E. Teufl, Asymptotics of the transition probabilities of the simple random walk on self-similar graphs. Trans. Am. Math. Soc. **356**, 393–414 (2004)
26. S. Kusuoka, A diffusion process on a fractal, in *Probabilistic Methods in Mathematical Physics (Katata/Kyoto, 1985)*, ed. by K. Itô, N. Ikeda (Academic, Boston, 1987), pp. 251–274
27. M.L. Lapidus, M. van Frankenhuysen, *Fractal Geometry and Number Theory, Complex Dimensions of Fractal Strings and Zeros of Zeta Functions* (Birkhäuser Boston, Boston, 2000)
28. G.M. Levin, M.L. Sodin, Polynomials with a disconnected Julia set, and the green mapping, in *Dynamical Systems and Complex Analysis* (Russian), ed. by V.A. Marchenko (Naukova Dumka, Kiev, 1992), pp. 17–24
29. T. Lindstrøm, *Brownian Motion on Nested Fractals*. Memoirs of the American Mathematical Society, vol. 420 (American Mathematical Society, Providence, 1990)
30. L. Malozemov, A. Teplyaev, Self-similarity, operators and dynamics. Math. Phys. Anal. Geom. **6**, 201–218 (2003)
31. V. Metz, Hilbert's projective metric on cones of Dirichlet forms. J. Funct. Anal. **127**(2), 438–455 (1995)
32. V. Metz, Renormalization contracts on nested fractals. J. Reine Angew. Math. **480**, 161–175 (1996)

33. V. Metz, The cone of diffusions on finitely ramified fractals. Nonlinear Anal. **55**(6), 723–738 (2003)
34. J. Milnor, *Dynamics in One Complex Variable*, 3rd edn. Annals of Mathematics Studies, vol. 160 (Princeton University Press, Princeton, 2006)
35. S. Minakshisundaram, Å. Pleijel, Some properties of the eigenfunctions of the Laplace-operator on Riemannian manifolds. Canad. J. Math. **1**, 242–256 (1949)
36. H. Poincaré, Sur une classe étendue de transcendantes uniformes. C. R. Acad. Sci. Paris **103**, 862–864 (1886)
37. H. Poincaré, Sur une classe nouvelle de transcendantes uniformes. J. Math. Pures Appl. IV. Ser. **6**, 316–365 (1890)
38. R. Rammal, G. Toulouse, Random walks on fractal structures and percolation clusters. J. Phys. Lett. **44**, L13–L22 (1983)
39. T. Ransford, *Potential Theory in the Complex Plane*. London Mathematical Society Student Texts, vol. 28 (Cambridge University Press, Cambridge, 1995)
40. S. Rosenberg, *The Laplacian on a Riemannian Manifold*. London Mathematical Society Student Texts, vol. 31 (Cambridge University Press, Cambridge, 1997)
41. C. Sabot, Existence and uniqueness of diffusions on finitely ramified self-similar fractals. Ann. Sci. École Norm. Sup. (4) **30**(5), 605–673 (1997)
42. T. Shima, The eigenvalue problem for the Laplacian on the Sierpiński gasket, in *Asymptotic Problems in Probability Theory: Stochastic Models and Diffusions on Fractals (Sanda/Kyoto, 1990)*, ed. by K.D. Elworthy, N. Ikeda. Pitman Research Notes in Mathematics Series, vol. 283 (Longman Scientific Technical, Harlow, 1993), pp. 279–288
43. T. Shima, On eigenvalue problems for Laplacians on p.c.f. self-similar sets. Jpn. J. Indust. Appl. Math. **13**, 1–23 (1996)
44. B.A. Steinhurst, A. Teplyaev, Existence of a meromorphic extension of spectral zeta functions on fractals. Lett. Math. Phys. **103**(12), 1377–1388 (2013)
45. R.S. Strichartz, Fractafolds based on the Sierpiński gasket and their spectra. Trans. Am. Math. Soc. **355**, 4019–4043 (2003)
46. R.S. Strichartz, Laplacians on fractals with spectral gaps have nicer Fourier series. Math. Res. Lett. **12**(2–3), 269–274 (2005)
47. R.S. Strichartz, A. Teplyaev, Spectral analysis on infinite Sierpiński fractafolds. J. Anal. Math. **116**, 255–297 (2012)
48. A. Teplyaev, Spectral zeta functions of fractals and the complex dynamics of polynomials. Trans. Am. Math. Soc. **359**(9), 4339–4358 (electronic) (2007)
49. G. Valiron, *Lectures on the General Theory of Integral Functions* (Private, Toulouse, 1923)
50. G. Valiron, *Fonctions Analytiques* (Presses Universitaires de France, Paris, 1954)
51. H. Weyl, Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung). Math. Ann. **71**, 441–479 (1912)

From Self-Similar Groups to Self-Similar Sets and Spectra

Rostislav Grigorchuk, Volodymyr Nekrashevych, and Zoran Šunić

To the memory of Gilbert Baumslag, a great colleague and a great friend

Abstract The survey presents developments in the theory of self-similar groups leading to applications to the study of fractal sets and graphs and their associated spectra.

Keywords Self-similar groups • Schreier graphs • Spectra • Self-similar sets • Fractals • Julia sets • Amenable action • Laplacian

Mathematics Subject Classification (2000). Primary 37A30; Secondary 28A80; 05C50, 20E08, 43A07

1 Introduction

The purpose of this survey is to present some recent developments in the theory of self-similar groups and its applications to the study of fractal sets. For brevity, we will concentrate only on the following two aspects (for other aspects see [7]):

(i) Construction of new fractals by using algebraic tools and interpretation of well known fractals (the first Julia set, Sierpiński gasket, Basilica fractal, and other Julia sets of post-critically finite rational maps on the Riemann sphere) in terms of self-similar groups and their associated objects – Schreier graphs.

The authors acknowledge partial support by the US National Science Foundation under Grants No. DMS-1105520 and DMS-1207699.

R. Grigorchuk (✉) • V. Nekrashevych • Z. Šunić
Department of Mathematics, Texas A&M University, College Station, TX 77843-3368, USA
e-mail: grigorch@math.tamu.edu; nekrash@math.tamu.edu; sunic@math.tamu.edu

C. Bandt et al. (eds.), *Fractal Geometry and Stochastics V*, Progress in Probability 70, DOI 10.1007/978-3-319-18660-3_11

(ii) Study of the spectra of the Laplacian on Schreier graphs of self-similar groups and on the associated fractals by appropriate limiting processes.

The presentation will be focused on a few representative examples for which the "entire program" (going from a self-similar group to its associated self-similar objects and calculation/description of their spectra) is successfully implemented, such as the first Grigorchuk group[1] $\mathcal{G}$, the lamplighter group $\mathcal{L}_2$, the 3-peg Hanoi Towers group $\mathcal{H}$, and the tangled odometers group $\mathcal{T}$, but also some examples with only partial implementation, such as the Basilica group $\mathcal{B}$ and the iterated monodromy group $\mathsf{IMG}\left(z^2+i\right)$.

2 Self-Similar Groups and Their Schreier Graphs

2.1 Schreier Graphs

Let G be a finitely generated group, generated by a finite symmetric set S (S being symmetric means $S = S^{-1}$) acting on a set Y (all actions in this survey will be left actions). The *Schreier graph* of the action of G on Y with respect to S is the oriented graph $\Gamma(G,S,Y)$ defined as follows. The vertex set of the Schreier graph is Y and the edge set is $S \times Y$. For $s \in S$ and $y \in Y$, the edge (s,y) connects y to sy. When the graph is drawn, the edge (s,y) is usually labeled just by s, since its orientation from y to sy uniquely indicates the correct "full label" (s,y). In other words, one usually draws $y \bullet \xrightarrow{s} \bullet\, sy$ instead of $y \bullet \xrightarrow{(s,y)} \bullet\, sy$.

The Schreier graph $\Gamma(G,S,Y)$ is connected if and only if the action is transitive (some authors define Schreier graphs only in the transitive/connected case).

Example 1 Let $Y = \{1,2,3,4\}$ and D_4 be the subgroup of the symmetric group on Y (with its usual left action) generated by $S = \langle \sigma, \bar{\sigma}, \tau \rangle$, where σ is the 4-cycle $\sigma = (1234)$, $\bar{\sigma}$ is its inverse $\bar{\sigma} = \sigma^{-1} = (1432)$, and τ is the transposition $\tau = (24)$ (note that one can interpret D_4 as the dihedral group of isometries of a square with vertices 1,2,3,4; σ is the rotation by $\pi/2$ and τ the mirror symmetry with respect to the line 13). The Schreier graph $\Gamma(D_4,S,Y)$ is drawn on the left in Fig. 1.

The edge (s,y) connects y to sy and the edge (s^{-1},sy) goes in the opposite direction and connects sy to y. In order to avoid clutter in the drawings, for each pair of mutually inverse generators $s, s^{-1} \in S$ that are not involutions, one usually chooses one of them, say s, and only draws the oriented edges labeled by s, while all edges labeled by s^{-1} are suppressed. Further, for an involution $s \in S$ and $y \in Y$, only one unoriented edge is drawn between y and sy (see the graph on the right in Fig. 1 and note that σ is not an involution, while τ is).

[1]The second and the third author insist on the use of this terminology.

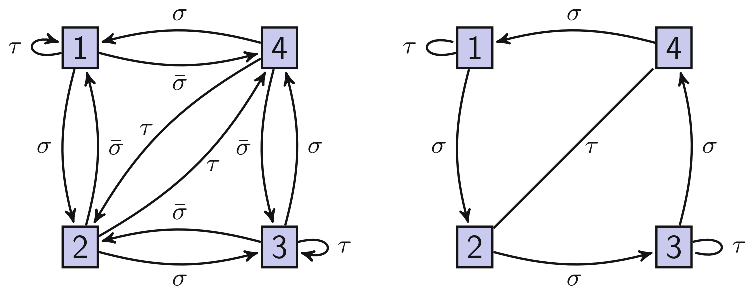


Fig. 1 The Schreier graph $\Gamma(D_4, S, Y)$, and its simplified drawing

2.2 Random Walk Operators on Schreier Graphs

The Schreier graph $\Gamma = \Gamma(G, S, Y)$ is regular with every vertex having both the out-degree and the in-degree equal to $|S|$. The *random walk* operator on Γ (also known as the *Markov operator*) is the operator

$$M : \ell^2(\Gamma) \to \ell^2(\Gamma)$$
$$(Mf)(y) = \frac{1}{|S|} \sum_{s \in S} f(sy).$$

where $\ell^2(\Gamma) = \ell^2(Y)$ is the Hilbert space of square summable functions on Y

$$\ell^2(\Gamma) = \ell^2(Y) = \left\{ f : Y \to \mathbb{R} \mid \sum_{y \in Y} |f(y)|^2 < \infty \right\}.$$

Thus, given a function $f : Y \to \mathbb{R}$ on the vertex set Y, the operator M produces an updated function $Mf : Y \to \mathbb{R}$ by replacing the value at each vertex y by the average of the f-values at the neighbors of y in the Schreier graph.

For $x \in \mathbb{R}$, let $M(x)$ be the operator $M(x) = M - xI$. The *spectrum* $\mathsf{Sp}(M)$ of M is the set of values of x for which the operator $M(x)$ from the pencil of operators $\{M(x) \mid x \in \mathbb{R}\}$ is not invertible. Note that the operator M is bounded (in fact $||M|| \leq 1$) and, since S is symmetric, it is self-adjoint. Therefore its spectrum is a closed subset of the interval $[-1, 1]$. When Y is finite, the spectrum $\mathsf{Sp}(M)$ is just the set of eigenvalues of the operator M, but in general the spectrum only contains the set of eigenvalues of M. Recall that λ is an *eigenvalue* of M if and only if $Mf = \lambda f$, for some nonzero function $f \in \ell^2(\Gamma)$; such a nonzero function is called an *eigenfunction* of M.

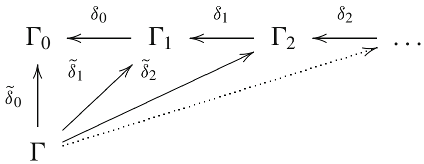
Let $G = \langle S \rangle$ act on two sets Y and $\tilde{Y}$ and $\delta : \tilde{Y} \to Y$ be a surjective G-equivariant map, that is, a surjective function δ such that $g\delta(\tilde{y}) = \delta(g\tilde{y})$, for $g \in G$ and $\tilde{y} \in \tilde{Y}$ (equivalently, $s\delta(\tilde{y}) = \delta(s\tilde{y})$, for $s \in S$ and $\tilde{y} \in \tilde{Y}$). On the level of Schreier graphs δ induces a surjective graph homomorphism from $\Gamma_{\tilde{Y}} = \Gamma(G, S, \tilde{Y})$ to $\Gamma_Y = \Gamma(G, S, Y)$ preserving edge labels and sending the edge $\tilde{y} \xrightarrow{s} s\tilde{y}$ to the edge $\delta(\tilde{y}) \xrightarrow{s} s\delta(\tilde{y})$. We say that $\Gamma_{\tilde{Y}}$ is a *covering* of Γ_Y and δ is a *covering map*.

Assume that both $\tilde{Y}$ and Y are finite. For every function $f \in \ell^2(\Gamma_Y)$, define the *lift* $\tilde{f} \in \ell^2(\Gamma_{\tilde{Y}})$ by $\tilde{f}(\tilde{y}) = f(\delta(\tilde{y}))$, for $\tilde{y} \in \tilde{Y}$. For all $f \in \ell^2(\Gamma_Y)$, we have

$$(M_{\tilde{Y}}\tilde{f})(\tilde{y}) = (M_Y f)(\delta(\tilde{y})).$$

If f is an eigenfunction of M_Y with eigenvalue λ, then $\tilde{f}$ is an eigenfunction of $M_{\tilde{Y}}$ with the same eigenvalue. Therefore, whenever there exists a surjective G-equivariant map $\delta : \tilde{Y} \to Y$ between two finite sets $\tilde{Y}$ and Y, the spectrum of M_Y is included in the spectrum of $M_{\tilde{Y}}$, that is, $\mathsf{Sp}(M_Y) \subseteq \mathsf{Sp}(M_{\tilde{Y}})$.

Let $\{Y_n\}_{n=0}^{\infty}$ be a sequence of finite G-sets (sets with a G-action defined on them), $\{\delta_n : Y_{n+1} \to Y_n\}_{n=0}^{\infty}$ a sequence of surjective G-equivariant maps, Y be a G-set, and $\{\tilde{\delta}_n : Y \to Y_n\}_{n=0}^{\infty}$ a sequence of surjective G-equivariant maps such that $\delta_n\tilde{\delta}_{n+1} = \tilde{\delta}_n$, for $n \geq 0$. Denote $\Gamma_n = \Gamma(G, S, Y_n)$, $\Gamma = \Gamma(G, S, Y)$, and the corresponding random walk operators by M_n and M, respectively. The sequences of equivariant maps $\{\delta_n\}$ and $\{\tilde{\delta}_n\}$ induce graph coverings between the corresponding Schreier graphs such that the following diagram commutes



(2.1)

and we obtain an increasing sequence $\{\mathsf{Sp}(M_n)\}_{n=0}^{\infty}$ of finite sets, each consisting of the eigenvalues of M_n. We are interested in situations in which this sequence is sufficient to determine the spectrum of M in the sense that

$$\overline{\bigcup_{n=0}^{\infty} \mathsf{Sp}(M_n)} = \mathsf{Sp}(M).$$

Example 2 This example is relatively straightforward, but it illustrates the setup we introduced above. Consider the infinite dihedral group $D_\infty = \langle a, b \rangle$, generated by two involutions a and b. We may think of it as the group of isometries of the set of integer points on the real line, with the action of a and b given by $a(n) = 1 - n$ and $b(n) = -n$. Let $Y = \mathbb{Z}$ and Γ be the Schreier graph $\Gamma = \Gamma(D_\infty, S, Y)$, drawn in the bottom row in Fig. 2. For $n \geq 0$, let $Y_n = \{0, \pm 1, \ldots, \pm 2^{n-1} - 1, 2^{n-1}\}$. Note that Y_n is a set of unique representatives of the residue classes modulo 2^n, for $n \geq 0$. Thus we may think of Y_n as $\mathbb{Z}/2^n\mathbb{Z}$. The action of D_∞ on $\mathbb{Z}$ induces a well defined action on the set of residue classes $\mathbb{Z}/2^n\mathbb{Z}$, for $n \geq 0$, and we denote $\Gamma_n = \Gamma(D_\infty, S, Y_n)$. The sequence of Schreier graphs $\{\Gamma_n\}$ is indicated in the top

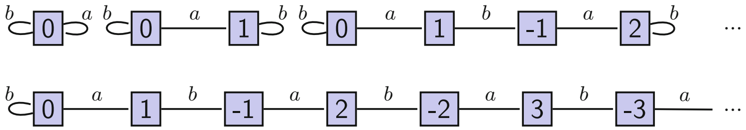


Fig. 2 The Schreier graphs $\Gamma_0, \Gamma_1, \Gamma_2, \ldots$ (*top row*) and Γ (*bottom row*) for the infinite dihedral group $D_\infty = \langle a, b \rangle$

row in Fig. 2. For $n \geq 0$, the maps $\delta_n : Y_{n+1} \to Y_n$ and $\tilde{\delta}_n : Y \to Y_n$, given by $\delta_n(y) = \mathsf{mod}(y, 2^n)$, for $y \in \mathbb{Z}/2^{n+1}\mathbb{Z}$, and $\tilde{\delta}_n(y) = \mathsf{mod}(y, 2^n)$, for $y \in \mathbb{Z}$, where $\mathsf{mod}(y, 2^n)$ is the remainder obtained when y is divided by 2^n, are D_∞-equivariant.

For $n \geq 0$, $\mathsf{Sp}(M_n)$ consists of 2^n distinct eigenvalues of multiplicity 1

$$\mathsf{Sp}(M_n) = \{1\} \cup \frac{1}{2}\bigcup_{i=0}^{n-1} f^{-i}(0) = \{1, 0\} \cup \frac{1}{2}\left\{ \underbrace{\pm\sqrt{2 \pm \sqrt{2 \pm \cdots \pm \sqrt{2}}}}_{i \text{ plus-minus signs}} \mid i=1,\ldots,n-1 \right\},$$

where $f(x) = x^2 - 2$. On the other hand, the spectrum of the doubly infinite path Γ is $[-1, 1]$ and we have

$$\mathsf{Sp}(M) = [-1, 1] = \frac{1}{2}\overline{\bigcup_{n=0}^{\infty} f^{-n}(0)} = \overline{\bigcup_{n=0}^{\infty} \mathsf{Sp}(M_n)}.$$

2.3 *Adjacency Operator on Schreier Graphs and Schreier Spectrum*

For the Schreier graph $\Gamma = \Gamma(G, S, Y)$ of the action of $G = \langle S \rangle$ on Y, the *adjacency operator* on Γ is the operator $A : \ell^2(\Gamma) \to \ell^2(\Gamma)$ defined by

$$(Af)(y) = \sum_{s \in S} f(sy).$$

The random walk operator $M = \frac{1}{|S|}A$ is the normalized version of the adjacency operator A and their spectra are just multiples of each other. Denote the spectrum of A by $\mathsf{Sp}(\Gamma)$ and call it the *Schreier spectrum* of Γ. This is the so called *adjacency spectrum*, but we want to emphasize the scope of all our considerations, namely, adjacency spectra of Schreier graphs of finitely generated groups. For the purposes of our calculations, the Schreier spectra turn out to be the most convenient, but it is easy to switch to their Markovian or Laplacian versions when needed (the *Laplacian* operator is the operator $L = I - M$, where I is the identity operator).

2.4 Rooted Regular Trees and Self-Similar Groups

We introduce the class of self-similar groups acting on regular rooted trees, providing a framework for examples like Example 2, and a source of other examples.

Let X be a finite set, usually called the *alphabet*, of size k. The set of all finite words over X is denoted by X^*. The set X^* can be naturally equipped with the structure of a *rooted k-regular tree* as follows. The vertices of the tree are the words in X^*, the *root* is the empty word ϵ, the *level n* is the set X^n of words of length n over X, and the children of each vertex $u \in X^*$ are the k vertices of the form ux, for $x \in X$. We use X^* to denote the set of finite words over X, the set of vertices of the rooted tree we just described, as well as the tree itself.

The group $\mathsf{Aut}(X^*)$ of all automorphisms of the rooted k-regular tree X^* preserves the root and all levels of the tree. Every automorphism $g \in \mathsf{Aut}(X^*)$ induces a permutation α_g of X, defined by $\alpha_g(x) = g(x)$, called the *root permutation* of g. It represents the action of g at the first letter in each word. For every automorphism $g \in \mathsf{Aut}(X^*)$ and every vertex $u \in X^*$, there exists a unique tree automorphism of X^*, denoted by g_u, such that, for all words $w \in X^*$,

$$g(uw) = g(u)g_u(w).$$

The automorphism g_u is called the *section* of g at u. It represents the action of g on the tails of words that start with u. Every automorphism g is uniquely determined by its root permutation α_g and the k sections at the first level g_x, for $x \in X$. Indeed, for every $x \in X$ and $w \in X^*$ we have

$$g(xw) = \alpha_g(x)g_x(w). \tag{2.2}$$

When $X = \{0, 1, \ldots, k-1\}$, a succinct representation, called *wreath recursion*, of the automorphism $g \in \mathsf{Aut}(X^*)$, describing its root permutation and its first level sections is given by

$$g = \alpha_g(g_0, g_1, \ldots, g_{k-1}). \tag{2.3}$$

In addition of being short and clear, it has many other advantages, not the least of which is that it emphasizes the fact that $\mathsf{Aut}(X^*)$ is isomorphic to the semidirect product $\mathsf{Sym}(X) \ltimes (\mathsf{Aut}(X^*))^X$, that is, to the permutational wreath product $\mathsf{Sym}(X) \wr_X \mathsf{Aut}(X^*)$, where $\mathsf{Sym}(X)$ is the group of all permutations of X.

A set $S \subseteq \mathsf{Aut}(X^*)$ of tree automorphisms is *self-similar* if it is closed under taking sections, that is, every section of every element of S is itself in the set S. Thus, for every word u, the action of every automorphism $s \in S$ on the tails of words that start with u looks exactly like the action of some element of S. Note that for a set S to be self-similar it is sufficient that it contains the first level sections of all of its elements. Indeed, this is because $g_{uv} = (g_u)_v$, for all words $u, v \in X^*$. A group $G \leq \mathsf{Aut}(X^*)$ of tree automorphisms is *self-similar* if it is self-similar as a set. Every group generated by a self-similar set is itself self-similar. This is because "sections of the product are products of sections" and "sections of the inverse are inverses of sections". To be precise, for all tree automorphisms g and h and all words $u \in X^*$,

$$(gh)_u = g_{h(u)} h_u \qquad \text{and} \qquad \left(g^{-1}\right)_u = \left(g_{g^{-1}(u)}\right)^{-1}$$

The observation that groups generated by self-similar sets are themselves self-similar enables one to easily construct many examples of finitely generated self-similar groups, as demonstrated in the next subsection.

Remark 1 It should be clarified that when we speak of a subset S or a subgroup G of $\mathsf{Aut}(X^*)$ as a self-similar set, we do not use this terminology in the, by now widely accepted and used, sense of Hutchinson [28]. It would be more precise to say, and it is often said, that the action is self-similar, that is, the action is adapted to the self-similar nature of the rooted tree and its boundary, the Cantor set. Self-similar sets in the sense of Hutchinson do play a role here, as such sets appear as limit spaces of contracting self-similar groups (see Sect. 3) and our considerations lead to results on Laplacians on such self-similar sets (see Sect. 7).

2.5 Automaton Groups

An *automaton*, in our context, is any finite self-similar set S of tree automorphisms. The group $G(S) = \langle S \rangle$, called the *automaton group* over S (or of S), is a finitely generated self-similar group. A simple way to define an automaton is by defining the action of each of its elements recursively as in (2.2).

Example 3 Consider the binary rooted tree based on the alphabet $X = \{0, 1\}^*$. Define a finite self-similar set $S = \{a, b\}$ of tree automorphisms recursively by

$$a(0u) = 1a(u) \qquad b(0u) = 0b(u),$$

$$a(1u) = 0b(u) \qquad b(1u) = 1a(u),$$

for every word $u \in X^*$, and $a(\epsilon) = b(\epsilon) = \epsilon$. Evidently, the root permutations and the sections of a and b are given in the following table.

s	α_s	s_0	s_1
a	(01)	a	b
b	$()$	b	a

where $()$ and (01) denote, respectively, the trivial and the nontrivial permutation of $X = \{0, 1\}$. Calculating the action of any element of S on any word in X^* by using the recursive definition is straightforward. For instance,

$$a(10101) = 0b(0101) = 00b(101) = 001a(01) = 0011a(1) = 00110.$$

One may think of the elements of an automaton S as the *states* of a certain type of transducer, a so-called *Mealy machine*. The recursive definition 2.2 of the action of $s \in S$ is interpreted as follows. To calculate the action of the state s on some input word xu starting with x, the machine first rewrites x into $\alpha_s(x)$, changes its state to s_x, and lets the new state handle the rest of the input u in the same manner. It reads the first letter of u, rewrites it appropriately, then moves to an appropriate state, which then handles the rest of the input, and so on, until the entire input word is read. It is common to represent the automaton S by an oriented labeled graph as follows. The vertex set is the set of states S, and each pair of a state $s \in S$ and a letter $x \in X$ determines a directed edge from s to s_x labeled by $x|\alpha_s(x)$ (equivalently, by $x|s(x)$).

Example 4 Four examples of finite self-similar sets of tree automorphisms are given in Fig. 3. The self-similar groups defined by these sets are the lamplighter group $\mathcal{L}_2 = \mathbb{Z} \ltimes (\oplus_{\mathbb{Z}} \mathbb{Z}/2\mathbb{Z})$ (top left), the dihedral group D_∞ (top right), the binary odometer group $\mathbb{Z}$ (bottom left), and the tangled odometers group $\mathcal{T}$ (bottom right). In the last three automata the state e represents the trivial automorphism of the tree, which does not change any input word. Thus, we use ϵ for the empty word, that is, the root of X^*, () for the trivial permutation of X, and e for the trivial automorphism of the tree X^*. To avoid clutter, in the automaton for $\mathbb{Z}$ we used the convention that the same edge may be used with several labels, while in the automaton for $\mathcal{T}$ the convention that the loops associated to the trivial state e are not drawn. Note that the first three automata are defined over the binary alphabet $X = \{0, 1\}$ while the last one is defined over the ternary alphabet $X = \{0, 1, 2\}$, hence that group acts on the ternary rooted tree.

One can easily switch back and forth between the various representations of the given automata. For instance, the recursive definition of the action of the dihedral group $D_\infty = \langle a, b\rangle$ on the binary rooted tree is given by

$$a(0u) = 1u, \qquad b(0u) = 0a(u),$$
$$a(1u) = 0u, \qquad b(1u) = 1b(u),$$

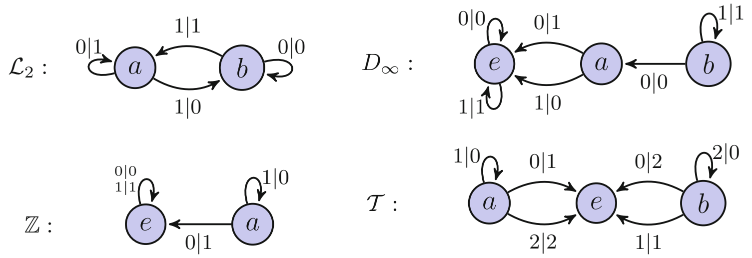


Fig. 3 Automata defining $\mathcal{L}_2$, D_∞, $\mathbb{Z}$, and $\mathcal{T}$

Tabular representation of the self-similar set defining $\mathcal{T}$ and the wreath recursion describing the same set are given on the left and on the right, respectively in

s	α_s	s_0	s_1	s_2
a	(01)	e	a	e
b	(02)	e	e	b

$$a = (01)(e, a, e) \qquad b = (02)(e, e, b) \tag{2.4}$$

It is clear that defining a finitely generated self-similar group is an easy task, in particular for automaton groups (note that not all finitely generated self-similar groups are automaton groups). One can methodically construct, one by one, all automaton groups by constructing all automata with a given number of states over an alphabet of a given size. However, it is not an easy task to recognize the group that is generated by a given automaton. A full classification of all automaton groups defined by automata with given number of states m and size of the alphabet k has been achieved only for $m = k = 2$ [15], while for the next smallest case $m = 3$ and $k = 2$ only a partial classification was obtained [6].

2.6 The Boundary Action and the Convergence $\Gamma_n \to \Gamma$

Let $G = \langle S \rangle$, with S symmetric and finite, be a finitely generated subgroup of $\mathsf{Aut}(X^*)$ and, for $n \geq 0$, let $\Gamma_n = \Gamma(G, S, X^n)$ be the corresponding Schreier graph of the action on level n. The map $\delta_n : X^{n+1} \to X^n$ given by deleting the last letter in each word is G-equivariant and induces a sequence of coverings of degree $|X|$

$$\Gamma_0 \xleftarrow{\delta_0} \Gamma_1 \xleftarrow{\delta_1} \Gamma_2 \xleftarrow{\delta_2} \ldots$$

Under the covering δ_n each of the $|X|$ edges $ux \bullet \xrightarrow{s} \bullet\, s(u)s_u(x)$ in Γ_{n+1}, for $x \in X$, is mapped to the edge $u \bullet \xrightarrow{s} \bullet\, s(u)$ in Γ_n.

Example 5 The *first Grigorchuk group* $\mathcal{G}$ is the self-similar group $\mathcal{G} = \langle a, b, c, d \rangle$ generated by four involutions a, b, c, and d acting on the binary tree and given by the wreath recursion

$$a = (01)(e, e), \qquad b = ()(a, c), \qquad c = ()(a, d), \qquad d = ()(e, b).$$

The Schreier graphs of its action on levels 0,1,2, and 3, are given in Fig. 4. This group was constructed by the first author in [17] as a particularly simple example of a finitely generated, infinite 2-group. It was the first example of a group of intermediate growth and the first example of an amenable group that is not elementary amenable [18] (we will get back to this aspect later).

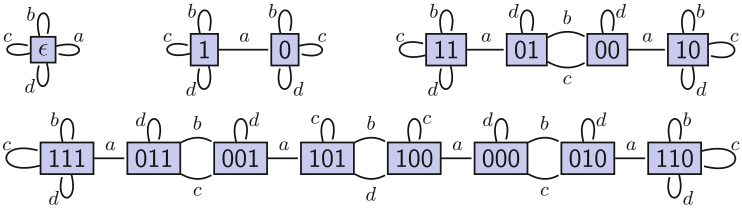


Fig. 4 The Schreier graphs Γ_0, Γ_1, Γ_2 (*top row*), and Γ_3 (*bottom row*) for the first Grigorchuk group $\mathcal{G}$

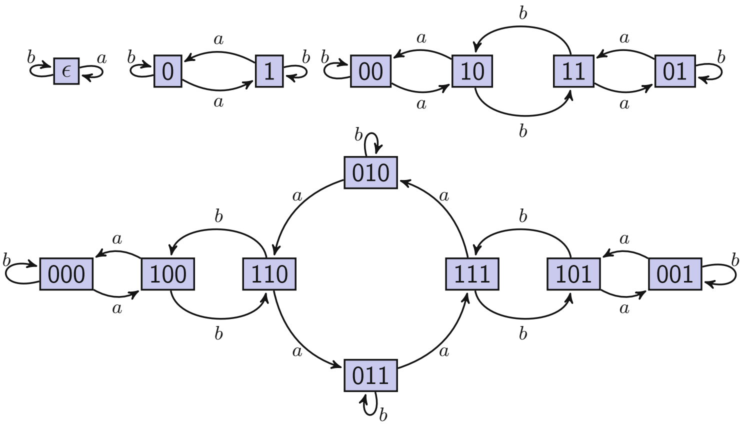


Fig. 5 The Schreier graphs Γ_0, Γ_1, Γ_2 (*top row*), and Γ_3 (*bottom row*) for the Basilica group $\mathcal{B}$

Example 6 The Basilica group is the self-similar group $\mathcal{B} = \langle a, b \rangle$ generated by the binary tree automorphisms a and b given by the wreath recursion

$$a = (01)(e, b), \qquad b = ()(e, a).$$

The Schreier graphs of its action on levels 0,1,2, and 3, are given in Fig. 5.

The group $\mathcal{B}$ was first considered in [26] and [27] where it was proved that it is a weakly branch, torsion free group which is not sub-exponentially amenable. It was later proved by Bartholdi and Virág [5], using speed estimates for random walks, that this group is amenable, thus providing the first example of an amenable group that is not sub-exponentially amenable.

Example 7 The Hanoi Towers group is the self-similar group $\mathcal{H} = \langle a, b, c \rangle$ generated by three involutions acting on the ternary tree given by the wreath recursion

$$a = (01)(e, e, a), \qquad b = (02)(e, b, e), \qquad c = (12)(c, e, e).$$

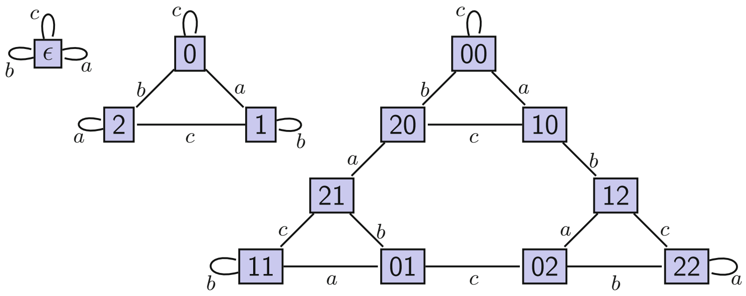


Fig. 6 The Schreier graphs Γ_0, Γ_1, and Γ_2 for the Hanoi Towers group $\mathcal{H}$

The Schreier graphs of its action on levels 0,1, and 2 are given in Fig. 6.

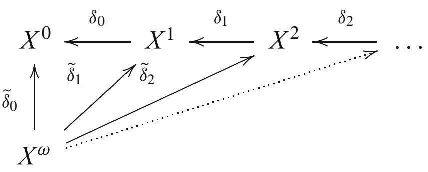
The group $\mathcal{H}$ was introduced in [21]. It models the well known Hanoi Towers game on three pegs in such a way that the Schreier graph Γ_n models the game for n disks. It is the first example of a finitely generated branch group that admits a surjective homomorphism onto the infinite dihedral group D_∞ (note that branch groups can only have virtually abelian proper quotients [19], and any finitely generated branch group that admits a surjective homomorphism to an infinite virtually abelian group must map onto $\mathbb{Z}$ or onto D_∞ [8]).

The *boundary* X^ω of the tree X^* is the space of ends of the tree X^*. More concretely, this is the space of all infinite rays

$$X^\omega = \{\, x_1x_2x_3\ldots \mid x_1, x_2, x_3, \dots \in X \,\},$$

that is, infinite paths without backtracking that start at the root. It has the structure of a metric space (in fact, ultrametric space) with metric defined by $d(\xi,\zeta) = 1/2^{|\xi\wedge\zeta|}$, where $\xi\wedge\zeta$ denotes the longest common prefix of the infinite rays ξ and ζ, and $|\xi\wedge\zeta|$ denotes its length. Thus, the longer the common prefix the closer the rays are. The induced topology is the product topology on $\prod_{i=1}^{\infty} X$, where the finite space X is given the discrete metric, implying that, topologically, the boundary X^ω is a Cantor set, and hence compact.

The action of any group of tree automorphisms $G \leq \mathsf{Aut}(X^*)$ naturally induces an action on the boundary of the tree X^*. The action of any automorphism $g \in \mathsf{Aut}(X^*)$ on X^ω is given by (2.2) with the understanding that w in that formula now applies to rays in X^ω, that is, to right-infinite words over X. If, for $n \geq 0$, we denote by $\tilde{\delta}_n : X^\omega \to X^n$ the map that deletes the tail of any ray beyond the first n-letters we obtain a sequence of G-equivariant maps. Thus we obtain the following diagram of G-equivariant maps.



(2.5)

Even if $G = \langle S \rangle$ acts *level transitively* on the tree X^* (transitively on each level X^n of the tree) and all Schreier graphs $\Gamma_n = \Gamma(G, S, X^n)$ are connected, the Schreier graph $\Gamma(G, S, X^\omega)$ of the action of G on the tree boundary is not connected. Indeed, since this graph is uncountable and the group G is countable, each orbit of the action on the boundary is countable and there must be uncountably many connected components (orbits) in the graph $\Gamma(G, S, X^\omega)$. Picking a connected component is equivalent to picking a point on the boundary that represents it, that is, picking an infinite ray $\xi = x_1x_2x_3 \cdots \in X^\omega$. Choose such a ray ξ and let $\Gamma = \Gamma_\xi = \Gamma(G, S, G\xi)$ be the Schreier graph of the boundary action of G on the orbit $G\xi = \{g(\xi) \mid g \in G\}$. We call the Schreier graph $\Gamma = \Gamma_\xi$ the *orbital Schreier graph* of G at ξ. It is a countable graph of degree $|S|$ and, since the restrictions of the maps $\tilde{\delta}_n$, for $n \geq 0$, to the orbit $G\xi$ are G-equivariant, the induced maps $\tilde{\delta}_n : \Gamma \to \Gamma_n$ are coverings. Therefore, we are precisely in the situation described by the diagram (2.1). Moreover, we can now state a sufficient condition under which the spectra of the sequence of finite graphs $\{\Gamma_n\}$ approximates the spectrum of Γ.

Theorem 2.1 (Bartholdi-Grigorchuk [3]) *Let $G = \langle S \rangle \leq \mathsf{Aut}(X^*)$ be a finitely generated, self-similar, level-transitive group of automorphisms of the rooted tree X^* and let $\xi \in X^\omega$ be a point on the tree boundary. For $n \geq 0$, let $\Gamma_n = \Gamma(G, S, X^n)$ be the Schreier graph of the action of G on level n of the tree and let $\Gamma = \Gamma_\xi = \Gamma(G, S, G\xi)$ be the orbital Schreier graph of G at ξ. If the action of G on the orbit $G\xi$ is amenable, then*

$$\overline{\bigcup_{n=0}^{\infty} \mathsf{Sp}(\Gamma_n)} = \mathsf{Sp}(\Gamma).$$

We recall the definition of an amenable action. The action of G on Y is *amenable* if there exists a normalized, finitely additive, G-invariant measure μ on all subsets of Y, that is, there exists a function $\mu : 2^Y \to [0, 1]$ such that

- (Normalization) $\mu(Y) = 1$,
- (Finite additivity) $\mu(A \sqcup B) = \mu(A) + \mu(B)$, for disjoint subsets $A, B \subseteq Y$,
- (G-invariance) $\mu(A) = \mu(gA)$, for $g \in G, A \subseteq Y$.

For a finitely generated group $G = \langle S \rangle$ (with S finite and symmetric, as usual) acting transitively on a set Y, the amenability of the action is equivalent to the amenability of the Schreier graph $\Gamma = \Gamma(G, S, Y)$ of the action and one of the many equivalent ways to define/characterize the amenability of Γ is as follows. The graph Γ is amenable if and only if

$$\inf\left\{ \frac{|\partial F|}{|F|} \mid F \text{ finite and nonempty set of vertices of } \Gamma \right\} = 0,$$

where the *boundary* ∂F of the set F is the set of vertices in Γ that are not in F but have a neighbor in F, that is, $\partial F = \{ v \in \Gamma \mid v \notin F \text{ and } sv \in F \text{ for some } s \in S \}$.

One sufficient condition for the amenability of the graph Γ is obtained by looking at its growth. Let Γ be any connected graph of uniformly bounded degree. Choose any vertex $v_0 \in \Gamma$ and, for $n \geq 0$, let $\gamma_{v_0}(n)$ be the number of vertices in Γ at combinatorial distance no greater than n from v_0. If the growth of $\gamma_{v_0}(n)$ is subexponential (that is, $\limsup_{n\to\infty} \sqrt[n]{\gamma_{v_0}(n)} = 1$), then Γ is an amenable graph.

By definition, a group G is amenable if its left regular action on itself is amenable. In such a case, every action of G is amenable and Theorem 2.1 applies. The class of amenable groups includes all finite and all solvable groups and is closed under taking subgroups, homomorphic images, extensions, and directed unions. The smallest class of groups that contains all finite and all abelian groups and is closed under taking subgroups, homomorphic images, extensions, and directed unions is known as the class of *elementary amenable* groups. There are amenable groups that are not elementary amenable and many such examples came from the theory of self-similar groups, starting with the first Grigorchuk group $\mathcal{G}$. The amenability of this group was proved by showing that it has subexponential (in fact intermediate, between polynomial and exponential) growth [18]. Other examples of amenable but not elementary amenable groups include Basilica group $\mathcal{B}$ [5], Hanoi Towers group $\mathcal{H}$, tangled odometers group $\mathcal{T}$, and many other automaton groups. See [4] and [1] for useful sufficient conditions for amenability of automaton groups based on random walk considerations and the notion of activity growth introduced by Sidki [41].

A large and interesting class of examples to which Theorem 2.1 applies is the class of contracting self-similar groups.

Definition 2.2 Let $G \leq \mathsf{Aut}(X^*)$ be a self-similar group of automorphisms of the rooted regular tree X^*. The group G is said to be *contracting* if there exists a finite set $\mathcal{N} \subseteq G$ such that, for every $g \in G$, there exists n such that $g_v \in \mathcal{N}$, for all words $v \in X^*$ of length at least n. The smallest set $\mathcal{N}$ satisfying this property is called the *nucleus* of the group.

Since the growth of each orbital Schreier graph Γ of a finitely generated, self-similar, contracting group is polynomial [3], such a graph Γ is amenable and, therefore, its spectrum can be approximated by the spectra of the finite graphs in the sequence $\{\Gamma_n\}$, as in Theorem 2.1. Note that it is not known yet whether all finitely generated contracting groups are amenable.

3 Iterated Monodromy Groups

The content of this section is not necessary in order to follow the rest of the survey, but it provides excellent examples, motivation, and context for our considerations.

3.1 Definition

Let $\mathcal{M}$ be a path connected and locally path connected topological space, and let $f : \mathcal{M}_1 \to \mathcal{M}$ be a finite degree covering map, where $\mathcal{M}_1$ is a subset of $\mathcal{M}$. The main examples for us are *post-critically finite complex rational functions*. Namely, a rational function $f \in \mathbb{C}(z)$ is said to be post-critically finite if the forward orbit $O_x = \{f^{\circ n}(x)\}_{n\geq 1}$ of every critical point x of f (seen as a self-map of the Riemann sphere $\hat{\mathbb{C}}$) is finite. Let P be the union of the forward orbits O_x, for all critical points. Denote $\mathcal{M} = \hat{\mathbb{C}} \setminus P$ and $\mathcal{M}_1 = f^{-1}(\mathcal{M})$. Then $\mathcal{M}_1 \subseteq \mathcal{M}$ and $f : \mathcal{M}_1 \to \mathcal{M}$ is a finite degree covering map.

Let $t \in \mathcal{M}$, and consider the *tree of preimages* T_f whose set of vertices is the disjoint union of the sets $f^{-n}(t)$, where $f^{-0}(t) = \{t\}$. We connect every vertex $v \in f^{-n}(t)$ to the vertex $f(v) \in f^{-(n-1)}(t)$. We then obtain a tree rooted at t.

If γ is a loop in $\mathcal{M}$ starting and ending at t then, for every $v \in f^{-n}(t)$, there exists a unique path γ_v starting at v such that $f^{\circ n} \circ \gamma_v = \gamma$. Denote by $\gamma(v)$ the end of the path γ_v. Then $v \mapsto \gamma(v)$ is an automorphism of the rooted tree T_f. We get in this way an action (called the *iterated monodromy action*) of the fundamental group $\pi_1(\mathcal{M}, t)$ on the rooted tree T_f. The quotient of the fundamental group by the kernel of the action is called the *iterated monodromy group* of f, and is denoted $\mathsf{IMG}\,(f)$. In other words, $\mathsf{IMG}\,(f)$ is the group of all automorphisms of T_f that are equal to a permutation of the form $v \mapsto \gamma(v)$ for some loop γ.

3.2 Computation of $\mathsf{IMG}\,(f)$

Let X be a finite alphabet of size $\deg f$, and let $\Lambda : X \to f^{-1}(t)$ be a bijection. For every $x \in X$, choose a path $\ell(x)$ starting at t and ending at $\Lambda(x)$. Let $\gamma \in \pi_1(\mathcal{M}, t)$. Denote by γ_x the path starting at $\Lambda(x)$ such that $f \circ \gamma_x = \gamma$, and let $\Lambda(y)$ be the end of γ_x. Then the paths $\ell(x)$, γ_x, and $\ell(y)^{-1}$ form a loop, which we will denote $\gamma|_x$ (see Fig. 7).

Proposition 3.1 (Nekrashevych [35]) *Let X be an alphabet in a bijection $\Lambda : X \longrightarrow f^{-1}(t)$. Let $\ell(x)$, y, and $\gamma|_x$ be as above. Then Λ can be extended to an*

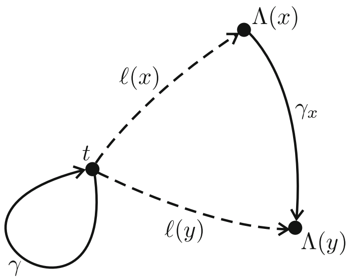


Fig. 7 Computation of IMG (f)

isomorphism of rooted trees $\Lambda : X^* \longrightarrow T_f$ *that conjugates the iterated monodromy action of* $\pi_1(\mathcal{M}, t)$ *on* T_f *with the action on* X^* *defined by the recursive rule:*

$$\gamma(xv) = y\gamma|_x(v).$$

In particular, IMG (f) *is a self-similar group.*

The self-similar action of IMG (f) on X^* described in the last proposition is called the *standard action*. It depends on the choice of the connecting paths $\ell(x)$, for $x \in X$, and the bijection $\Lambda : X \to f^{-1}(t)$. Changing the connecting paths amounts to post-composition of the wreath recursion with an inner automorphism of the wreath product $\mathsf{Sym}(X) \wr_X$ IMG (f).

Example 8 (Basilica group $\mathcal{B}$ = IMG $(z^2 - 1)$*)* The polynomial $z^2 - 1$ is post-critically finite with $P = \{0, -1, \infty\}$. The fundamental group of $\hat{\mathbb{C}} \setminus P$ is generated by two loops a, b going around the punctures 0 and -1, respectively. With an appropriate choice of the connecting paths (see [35, Subsection 5.2.2.]), the wreath recursion for IMG $(z^2 - 1)$ is exactly the same as the one in Example 6. Thus, $\mathcal{B}$ = IMG $(z^2 - 1)$.

Example 9 (Tangled odometers group $\mathcal{T}$ = IMG $\left(-\frac{z^3}{2} + \frac{3z}{2}\right)$*)* The polynomial $f(z) = -z^3/2 + 3z/2$ has three critical points: 1, -1, and ∞. All of them are fixed points of f, hence $P = \{1, -1, \infty\}$, and the fundamental group of is generated by loops around 1 and -1. The corresponding iterated monodromy group is defined by the wreath recursion (2.4), and this is the tangled odometers group $\mathcal{T}$.

Example 10 (Hanoi Towers group $\mathcal{H}$ = IMG $\left(z^2 - \frac{16}{27z}\right)$*)* The iterated monodromy group of the rational function $z^2 - 16/(27z)$ is conjugate in $\mathsf{Aut}(X^*)$ to the Hanoi Towers group $\mathcal{H}$ (see [22]).

Example 11 (Dihedral group D_∞ = IMG $(z^2 - 2)$ *and binary odometer group* $\mathbb{Z}$ = IMG (z^2)*)* The iterated monodromy group of the polynomial $z^2 - 2$ is the dihedral group D_∞ and of the polynomial z^2 is the binary odometer group $\mathbb{Z}$ (infinite cyclic group) from Example 4.

3.3 Limit Spaces of Contracting Self-Similar Groups

Suppose that G is a contracting self-similar group. Let $X^{-\omega}$ be the space of all left-infinite sequences $\ldots x_2x_1$ of elements of X with the direct product topology. We say that two sequences $\ldots x_2x_1$ and $\ldots y_2y_1$ in $X^{-\omega}$ are *asymptotically equivalent* if there exists a sequence $\{g_k\}_{k=1}^{\infty}$ of elements in G, taking a finite set of values, such that $g_k(x_k \ldots x_1) = y_k \ldots y_1$, for all $k \geq 1$. It is easy to see that this is an equivalence relation. The *limit space* of G is the quotient of the topological space $X^{-\omega}$ by the asymptotic equivalence relation. It is always a metrizable space of finite topological dimension (if G is contracting). Note that the asymptotic equivalence relation is invariant with respect to the shift $\ldots x_2x_1 \mapsto \ldots x_3x_2$. Consequently, the shift induces a continuous self-map on the limit space of G. The obtained map is called the *limit dynamical system* of the group G.

Theorem 3.2 (Nekrashevych [35]) *Suppose that f is a post-critically finite complex rational function. Then $\mathsf{IMG}\,(f)$ is a contracting self-similar group with respect to any standard action. The limit dynamical system of $\mathsf{IMG}\,(f)$ is topologically conjugate to the restriction of f onto its Julia set.*

The *Julia set* of a complex rational function f can be defined as the closure of the set of points c such that there exists n such that $f^n(c) = c$ and $\left| \frac{d}{dz} f^n(z)\Big|_{z=c} \right| > 1$. The Julia sets of $z \mapsto z^2 - 1$, $z \mapsto -\frac{z^3}{2} + \frac{3z}{2}$, and $z \mapsto z^2 - \frac{16}{27z}$ are given in Fig. 8. Theorem 3.2 provides context and explanation for the striking similarity between the structure of the Schreier graphs of the Basilica group in Fig. 5 and the Basilica fractal in Fig. 8, as well as between the structure of the Schreier graphs of the Hanoi Towers group in Fig. 6 and the Sierpiński gasket in Fig. 8.

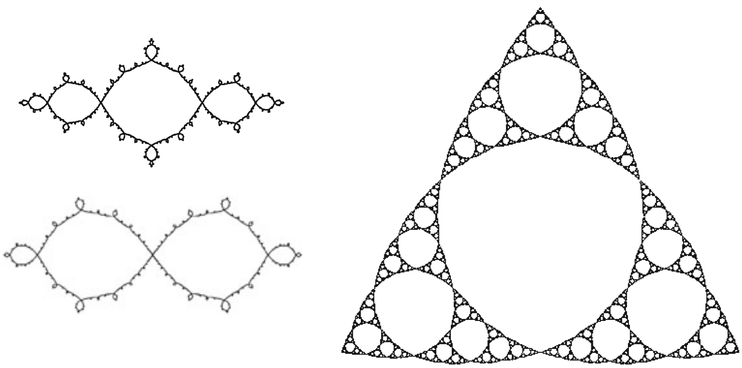


Fig. 8 Julia set of $z \mapsto z^2 - 1$ (*top left*), $z \mapsto -\frac{z^3}{2} + \frac{3z}{2}$ (*bottom left*), and $z \mapsto z^2 - \frac{16}{27z}$ (*right*)

4 Relation to Other Operators and Spectra

4.1 Hecke Type Operators

Let $G = \langle S \rangle$, with S finite and symmetric, be a finitely generated group and $\lambda : G \to \mathcal{U}(\mathcal{W})$ a unitary representation of G on a Hilbert space $\mathcal{W}$. To each element $\mathfrak{m} = \sum_{i=1}^{n} \alpha_i \cdot g_i$ of the group algebra $\mathbb{C}[G]$ one can associate the operator

$$\lambda(\mathfrak{m}) = \sum_{i=1}^{n} \alpha_i \lambda(g_i).$$

In particular, we consider the *Hecke type operator* H_λ on the Hilbert space $\mathcal{W}$ associated to the group algebra element $\mathfrak{h} = \frac{1}{|S|} \sum_{s \in S} s$ and given by

$$H_\lambda = \frac{1}{|S|} \sum_{s \in S} \lambda(s).$$

4.2 Koopman Representation and Hecke Type Operators

Let G be a countable group acting on a measure space (Y, μ) by measure-preserving transformations. The *Koopman representation* π is the unitary representation of G on the Hilbert space $L^2(Y, \mu)$ given by

$$(\pi(g)f)(y) = f(g^{-1}y)$$

for $f \in L^2(Y, \mu)$ and $y \in Y$.

Let $G = \langle S \rangle \leq \mathsf{Aut}(X^*)$ be a finitely generated, self-similar, level-transitive group of automorphisms of the rooted regular tree X^*. Note that the boundary X^ω, which has the structure of a Cantor set $\prod_{i=1}^{\infty} X$, is a measure space with respect to the product of uniform measures on X (for the cylindrical set uX^*, we have $\mu(uX^*) = \frac{1}{|X|^{|u|}}$). The group G acts on X^ω by measure-preserving transformations and we may consider the Koopman representation π of G on $L^2(X^\omega, \mu)$ and the associated Hecke type operator H_π on $L^2(X^\omega, \mu)$, given by

$$H_\pi = \frac{1}{|S|} \sum_{s \in S} \pi(s).$$

For every $n \geq 0$, we may also consider the representation π_n on $L^2(X^n, \mu_n)$ on the finite probability space X^n with uniform probability measure μ_n, corresponding to level n of the tree, and the associated Hecke type operator

$$H_{\pi_n} = \frac{1}{|S|} \sum_{s \in S} \pi_n(s).$$

Denote $\mathsf{Sp}(H_\pi) = \mathsf{Sp}(\pi)$ and $\mathsf{Sp}(H_{\pi_n}) = \mathsf{Sp}(\pi_n)$, for $n \geq 0$.

Theorem 4.1 (Bartholdi-Grigorchuk [3]) *Let G be a finitely generated, self-similar, level-transitive group of automorphisms of the rooted regular tree X^*. Then*

$$\mathsf{Sp}(\pi) = \overline{\bigcup_{n=0}^{\infty} \mathsf{Sp}(\pi_n)}.$$

Note that, unlike in Theorem 2.1, no additional requirements (such as amenability of the action) are needed in the last result.

4.3 Quasi-Regular Representations and Hecke Type Operators

It is well known that every transitive left action of a group G on any set Y is equivalent to the action of G on the left coset space G/P, where $P = \mathsf{Stab}_G(y)$ is the stabilizer of the point $y \in Y$ (since the action is transitive this point may be chosen arbitrarily). In fact, Schreier graphs originate as the graphs of the action of groups on their coset spaces.

For a countable group G and any subgroup $P \leq G$, the *quasi-regular representation* is the unitary representation $\rho_{G/P}$ of G on the Hilbert space $\ell^2(G/P)$ given by

$$(\rho_{G/P}(g)f)(hP) = f(g^{-1}hP),$$

for $f \in \ell^2(G/P)$ and $h \in G$. When P is the trivial group we obtain the *left regular representation* ρ_G defined by

$$(\rho_G(g)f)(h) = f(g^{-1}h),$$

for $f \in \ell^2(G)$ and $h \in G$.

Let $G = \langle S \rangle \leq \mathsf{Aut}(X^*)$ be a finitely generated, self-similar, level-transitive group of automorphisms of the rooted regular tree X^* and let $\xi = x_1x_2x_3\ldots$ be a point on the boundary X^ω. For $n \geq 0$, the point $x_1x_2\ldots x_n$ is the unique point at level n on the ray ξ. Let

$$P_n = \mathsf{Stab}_G(x_1x_2\ldots x_n), \text{ for } n \geq 0, \text{and}$$

$$P = \mathsf{Stab}_G(\xi).$$

Note that $\bigcap_{n=0}^{\infty} P_n = P_\xi$.

Denote by ρ_n the quasi-regular representation ρ_{G/P_n} corresponding to the subgroup P_n (thus, to the action of G on level n of the tree) and by ρ_ξ the representation ρ_{G/P_ξ}. We consider the Hecke type operator H_{ρ_ξ} on $\ell^2(G/P_\xi)$

$$H_{\rho_\xi} = \frac{1}{|S|}\sum_{s\in S}\rho_\xi(s)$$

and, for $n \geq 0$, the Hecke type operator

$$H_{\rho_n} = \frac{1}{|S|}\sum_{s\in S}\rho_n(s).$$

Denote $\mathsf{Sp}(H_{\rho_\xi}) = \mathsf{Sp}(\rho_\xi)$ and $\mathsf{Sp}(H_{\rho_n}) = \mathsf{Sp}(\rho_n)$, for $n \geq 0$.

The following result extends Theorem 2.1 and compares the Schreier spectrum to the spectrum of the Hecke type operators H_π and H_{ρ_ξ} associated to the Koopman representation π and the quasi-regular representation ρ_ξ, respectively.

Theorem 4.2 (Bartholdi-Grigorchuk [3])

(a) Let $G = \langle S\rangle \leq \mathsf{Aut}(X^*)$ be a finitely generated, self-similar, level-transitive group of automorphisms of the rooted regular tree X^* and let $\xi \in X^\omega$. Then, for $n \geq 0$,

$$\frac{1}{|S|}\mathsf{Sp}(\Gamma_n) = \mathsf{Sp}(\rho_n) = \mathsf{Sp}(\pi_n)$$

and

$$\frac{1}{|S|}\mathsf{Sp}(\Gamma_\xi) = \mathsf{Sp}(\rho_\xi) \subseteq \mathsf{Sp}(\pi).$$

(b) If the action of G on the orbit $G\xi$ is amenable, then

$$\frac{1}{|S|}\overline{\bigcup_{n=0}^{\infty}\mathsf{Sp}(\Gamma_n)} = \frac{1}{|S|}\mathsf{Sp}(\Gamma_\xi) = \mathsf{Sp}(\rho_\xi) = \mathsf{Sp}(\pi).$$

(c) If the group P_ξ is amenable, then

$$\frac{1}{|S|}\mathsf{Sp}(\Gamma_\xi) = \mathsf{Sp}(\rho_\xi) \subseteq \mathsf{Sp}(\rho_G),$$

where ρ_G is the left-regular representation of G (and $\mathsf{Sp}(\rho_G)$ is the spectrum of the corresponding Hecke type operator H_{ρ_G}).

By part (b) in the last result, if the group G is amenable, then all orbital Schreier graphs have the same spectrum (there is no dependence on the choice of the point

$\xi \in X^\omega$, since the representation π does not depend on it). More generally, if all orbital Schreier graphs Γ_ξ, for $\xi \in X^\omega$ are amenable, as it happens in the case of contracting self-similar groups, then they all have the same spectrum. Examples of nonamenable groups with amenable orbital Schreier graphs Γ_ξ were provided in [13] (thus, part (b) applies to some nonamenable groups).

We point out that part (b) is mistakenly stated in [3] under the assumption that either the action of G on the orbit $G\xi$ is amenable or P_ξ is amenable. The assumption that P_ξ is amenable only applies in part (c), and this part of Theorem 4.2 follows from [3, Proposition 3.5].

5 Method of Computation

The method of computation of spectra, introduced in [3] and further implemented and refined in [14, 16, 23, 25] is based on the use of invariant sets of multidimensional rational maps and the Schur complement. We will present the approach in the next two subsections, one addressing the global picture, and the other the details.

5.1 A Global Preview of the Method

Let A be an operator for which we would like to calculate the spectrum. Include A and the entire pencil $\{A(x) \mid x \in \mathbb{C}\}$ with $A(x) = A - xI$ into a multidimensional pencil of operators

$$\{ A^{(d)}(x_1, x_2, \ldots, x_d) \mid x_1, \ldots, x_d \in \mathbb{C} \}$$

such that

$$A(x) = A^{(d)}(x, x_2^{(0)}, x_3^{(0)}, \ldots, x_d^{(0)}),$$

for some particular values $x_2^{(0)}, x_3^{(0)}, \ldots, x_d^{(0)} \in \mathbb{C}$. Define the joint spectrum by

$$\mathsf{Sp}(A^{(d)}) = \left\{ (x_1, x_2, \ldots, x_d) \in \mathbb{C}^d \;\middle|\; A^{(d)}(x_1, x_2, \ldots, x_d) \text{ is not invertible } \right\}.$$

Then

$$\mathsf{Sp}(A) = \mathsf{Sp}(A^{(d)}) \cap \ell,$$

where ℓ is the line

$$\ell = \left\{ (x_1, x_2, \ldots, x_d) \in \mathbb{C}^d \;\middle|\; x_2 = x_2^{(0)}, x_3 = x_3^{(0)}, \ldots, x_d = x_d^{(0)} \right\}$$

in the d-dimensional space $\mathbb{C}^d$.

In the case of a self-adjoint operator A, which is always our case, we can use the field $\mathbb{R}$ instead of $\mathbb{C}$.

The problem naturally splits into three steps:

(i) Determine a suitable higher-dimensional pencil containing $\{A(x) \mid x \in \mathbb{R}\}$.
(ii) Determine the joint spectrum $\mathsf{Sp}(A^{(d)})$.
(iii) Determine the intersection $\mathsf{Sp}(A) = \mathsf{Sp}(A^{(d)}) \cap \ell$.

In the examples that were successfully treated by this approach, the joint spectrum $\mathsf{Sp}(A^{(d)})$ is an invariant set under some rational d-dimensional map $F : \mathbb{R}^d \to \mathbb{R}^d$. Thus, in practice, the step (ii) is understood as

(ii)′ Determine the joint spectrum $\mathsf{Sp}(A^{(d)})$ as an F-invariant set for a suitable d-dimensional rational map $F : \mathbb{R}^d \to \mathbb{R}^d$.

It may be somewhat counterintuitive why one should "increase the dimension of the problem in order to solve it", but the method has worked well in situations were direct approaches have failed. What happens is that the joint spectrum in $\mathbb{R}^d$, corresponding to the d-fold pencil of operators, is sometimes well behaved and easier to describe than the spectrum of the original 1-fold pencil. On the other hand, even when appropriate $A^{(d)}$ and F are found, the structure of the F-invariant set can be quite complicated and have the shape of a "strange attractor".

5.2 More Details

Let $G = \langle S \rangle$ be an automaton group generated by the elements of the finite and symmetric self-similar set S. For $n \geq 0$, the representations π_n and ρ_n are equivalent and may be viewed as representations on the $|X|^n$-dimensional vector space $\ell^2(X^n)$. The $|X|^n \times |X|^n$ adjacency matrix A_n (the rows and the columns are indexed by the words over X of length n) of Γ_n is given by

$$A_n = \sum_{s \in S} \pi_n(s).$$

The $|X|^n \times |X|^n$ matrix $\pi_n(s)$ is given recursively, for $n > 0$, by blocks of size $|X|^{n-1} \times |X|^{n-1}$

$$\pi_n(s) = \left[B_{y,x}(s)\right]_{y,x \in X} \tag{5.1}$$

corresponding to the decomposition

$$\ell^2(X^n) = \bigoplus_{x \in X} \ell^2(xX^{n-1}),$$

and the block $B_{y,x}(s)$ is given by

$$B_{yx}(s) = \begin{cases} \pi_{n-1}(s_x), & s(x) = y \\ 0, & \text{otherwise} \end{cases}$$

For $n = 0$, the space $\ell^2(X^0)$ corresponding to the root of the tree is 1-dimensional and $\pi_0(s)$ is the 1×1 identity matrix $\pi_0(s) = [1]$. We call (5.1) the *matrix wreath recursion* of S (it directly corresponds to the wreath recursion that defines the generators $s \in S$).

From now on, we use the notation $s_n = \pi_n(s)$.

Example 12 For the first Grigorchuk group $\mathcal{G}$ the matrix wreath recursion gives

$$a_0 = b_0 = c_0 = d_0 = [1]$$

and for $n > 0$,

$$a_n = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \qquad b_n = \begin{bmatrix} a_{n-1} & 0 \\ 0 & c_{n-1} \end{bmatrix} \qquad c_n = \begin{bmatrix} a_{n-1} & 0 \\ 0 & d_{n-1} \end{bmatrix} \qquad d_n = \begin{bmatrix} 1 & 0 \\ 0 & b_{n-1} \end{bmatrix},$$

where, in each case, 0 and 1 denote the zero matrix and the identity matrix, respectively, of appropriate size ($2^{n-1} \times 2^{n-1}$). Therefore, $A_0 = [4]$ and, for $n > 0$,

$$A_n = \begin{bmatrix} 2a_{n-1} + 1 & 1 \\ 1 & b_{n-1} + c_{n-1} + d_{n-1} \end{bmatrix}.$$

Example 13 For the tangled odometers group $\mathcal{T}$ the matrix wreath recursion gives

$$a_0 = b_0 = a_0^{-1} = b_0^{-1} = [1]$$

and for $n \geq 0$,

$$a_{n+1} = \begin{bmatrix} 0 & a_n & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad (a^{-1})_{n+1} = \begin{bmatrix} 0 & 1 & 0 \\ (a^{-1})_n & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$b_{n+1} = \begin{bmatrix} 0 & 0 & b_{n-1} \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \qquad (b^{-1})_{n+1} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ (b^{-1})_n & 0 & 0 \end{bmatrix}.$$

Therefore, $A_0 = [4]$ and, for $n \geq 0$,

$$A_{n+1} = \begin{bmatrix} 0 & 1 + a_n & 1 + b_n \\ 1 + (a^{-1})_n & 2 & 0 \\ 1 + (b^{-1})_n & 0 & 2 \end{bmatrix}.$$

Once the recursive definition of the adjacency operator A_n is established we consider the matrix

$$A_n(x) = A_n - xI = \left(\sum_{s \in S} s_n \right) - xI,$$

and more generally, a matrix of the form

$$A_n^{(d)}(x_1, \ldots, x_d) = A_n - x_1 I - \left(\sum_{i=2}^{d} x_i \cdot g_i \right) = \left(\sum_{s \in S} s_n \right) - x_1 I - \left(\sum_{i=2}^{d} x_i \cdot g_i \right),$$

for some auxiliary operators $g_2, \ldots, g_d$. There is no known general approach how to choose appropriate auxiliary operators. In practice, one needs to come up with good choices that make the subsequent calculations feasible.

We then calculate, by using elementary column and row transformations and the Schur complement, the determinant of $A_n^{(d)}$ in terms of the determinant of $A_{n-1}^{(d)}$ and obtain a recursive expression of the form

$$\det(A_n^{(d)}(x_1, \ldots, x_d)) = P_n(x_1, \ldots, x_d) \det(A_{n-1}^{(d)}(F(x_1, \ldots, x_d))), \tag{5.2}$$

where $P_n(x_1, \ldots, x_d)$ is a polynomial function and $F : \mathbb{R}^d \to \mathbb{R}^d$ is a rational function in the variables $x_1, \ldots, x_d$. Clearly, if the point $(x_1', \ldots, x_d')$ is in the zero set of $\det(A_{n-1}^{(d)}(x_1, \ldots, x_d))$, then any point in $F^{-1}(x_1', \ldots, x_d')$ is in the zero set of $\det(A_n^{(d)}(x_1, \ldots, x_d))$. Thus, describing the joint spectrum through iterations of the recursion (5.2) leads to iterations of the rational map F.

Understanding the structure of the zero sets of $\det(A_n^{(d)}(x_1, \ldots, x_d))$, for $n \geq 0$, and relating them to the zero sets of $\det(A_n(x))$ is accomplished, in the situations when we are able to resolve this problem, by finding a function $\psi : \mathbb{R}^d \to \mathbb{R}$ and a polynomial function $f : \mathbb{R} \to \mathbb{R}$ such that

$$\psi(F(x_1, \ldots, x_d)) = f(\psi(x_1, \ldots, x_d)),$$

that is, by finding a semi-conjugacy from the d-dimensional rational function F to a polynomial function f in a single variable. Since we have

$$\psi(F^{\circ m}(x_1, \ldots, x_d)) = f^{\circ m}(\psi(x_1, \ldots, x_d)),$$

the iterations of F are related to the iterations of f and then the desired spectrum is described through the iterations of the latter.

6 Concrete Examples and Computation Results

In this section we present several concrete examples of calculations of spectra based on the method suggested in the Sect. 5. All groups in this section are amenable. By Theorem 4.2, the choice of the point on the boundary is irrelevant for the Schreier spectrum and this is why no such choice is discussed in these examples.

One of the examples, the Hanoi Towers group $\mathcal{H}$, leads to results on the Sierpiński gasket. The spectrum of Sierpiński gasket goes back to the work of the physicists Rammal and Toulouse [38]. It was turned into a mathematical framework by Fukushima and Shima [9]. Note that, in these works, the Sierpiński gasket was approximated by a sequence of graphs that are 4-regular (with the exception of the three corner vertices, which have degree 2), while our approach yields an approximation through a different, but related, sequence of 3-regular graphs. A method for spectra calculations in more general cases, called *spectral decimation*, was developed by Kumagai, Malozemov, Shima, Teplyaev, Strichartz and others [31, 33, 34, 40, 42, 43]. Connections with Julia sets are well-known, as for instance given by Teplyaev [44].

6.1 The First Grigorchuk Group $\mathcal{G}$

As was already mentioned, the method sketched above was introduced in [3] in order to compute the spectrum of the sequence of Schreier graphs $\{\Gamma_n\}$ and the boundary Schreier graph Γ for the case of the first Grigorchuk group $\mathcal{G}$, as well as several other examples, including the Gupta-Sidki 3-group [20].

Theorem 6.1 (Bartholdi-Grigorchuk [3]) *For $n \geq 1$, the spectrum of the graph Γ_n, as a set, has 2^n elements (thus, all eigenvalues are distinct) and is equal to*

$$\mathsf{Sp}(\Gamma_n) = \left\{ 1 \pm \sqrt{5 + 4\cos\frac{2k\pi}{2^n}} \;\mid\; k = 0, \ldots, 2^{n-1} \right\} \setminus \{-2, 0\}.$$

The spectrum of Γ (the Schreier spectrum of $\mathcal{G}$), as a set, is equal to

$$\mathsf{Sp}(\Gamma) = [-2, 0] \cup [2, 4].$$

Remark 2 There is a different way in which the spectrum of Γ_n can be written. Namely, for $n \geq 2$,

$$\mathsf{Sp}(\Gamma_n) = \{4, 2\} \cup \left(1 \pm \sqrt{5 \pm 2 \bigcup_{i=0}^{n-2} f^{-i}(0)} \right),$$

where

$$f(x) = x^2 - 2.$$

Note that

$$f^{-k}(0) = \pm\sqrt{2 \pm \sqrt{2 \pm \sqrt{2 \pm \cdots \pm \sqrt{2}}}},$$

where the root sign appears exactly k times. The closure $\overline{\bigcup_{i=0}^{\infty} f^{-i}(0)}$ is equal to the interval $[-2, 2]$ and is the Julia set of the polynomial f, Therefore,

$$\begin{aligned}\mathsf{Sp}(\Gamma) &= \{4, 2\} \cup \left(1 \pm \sqrt{5 \pm 2 \cdot [-2, 2]}\right) = \{4, 2\} \cup \left(1 \pm \sqrt{[1, 9]}\right) \\ &= \{4, 2\} \cup (1 \pm [1, 3]) = [-2, 0] \cup [2, 4].\end{aligned}$$

For the calculations in this example, we may use the 2-dimensional auxiliary pencil of operators defined by

$$A_n^{(2)}(x, y) = a_n + b_n + c_n + d_n - (1 + x)I + (y - 1)a_n.$$

The recursive formula for the determinant of $A_n(x, y)$ is, for $n \geq 2$,

$$\det(A_n^{(2)}(x, y)) = \left(x^2 - 4\right)^{2^{n-2}} \det(A_{n-1}^{(2)}(F(x, y))),$$

where $F : \mathbb{R}^2 \to \mathbb{R}^2$ is given by

$$F(x, y) = \left(x - \frac{xy^2}{x^2 - 4}, \frac{2y^2}{x^2 - 4}\right).$$

The map $\psi : \mathbb{R}^2 \to \mathbb{R}$ that semi-conjugates F to $f(x) = x^2 - 2$ is

$$\psi(x, y) = \frac{x^2 - 4 - y^2}{2y}.$$

The 2-dimensional joint spectrum of $A_n(x, y)$ is a family of hyperbolae and intersecting this family with the line $y = 1$ gives the desired spectrum.

The more general problem of determining the spectrum of the operator associated to any element of the form $ta + ub + vc + wd$ in the group algebra $\mathbb{R}[\mathcal{G}]$ is considered in [11], where it is shown that, apart from few exceptions (such as the case $u = v = w$ considered above), the spectrum is always a Cantor set.

6.2 The Hanoi Towers Group $\mathcal{H} = \mathsf{IMG}\left(z^2 - \frac{16}{27z}\right)$ and Sierpiński Gasket

Theorem 6.2 (Grigorchuk-Šunić [21, 23]) *For $n \geq 1$, the spectrum of the graph Γ_n, as a set, has $3 \cdot 2^{n-1} - 1$ elements and is equal to*

$$\{3\} \cup \bigcup_{i=0}^{n-1} f^{-i}(0) \cup \bigcup_{j=0}^{n-2} f^{-j}(-2),$$

where

$$f(x) = x^2 - x - 3.$$

The multiplicity of the 2^i eigenvalues in $f^{-i}(0)$, $i = 0, \dots, n-1$, is a_{n-i}, and the multiplicity of the 2^j eigenvalues in $f^{-j}(-2)$, $j = 0, \dots, n-2$, is b_{n-j}, where, for $m \geq 1$,

$$a_m = \frac{3^{m-1} + 3}{2} \qquad \text{and} \qquad b_m = \frac{3^{m-1} - 1}{2}.$$

The spectrum of Γ (the Schreier spectrum of $\mathcal{H}$), as a set, is equal to

$$\overline{\bigcup_{i=0}^{\infty} f^{-i}(0)}.$$

It consists of a set of isolated points, the backward orbit $I = \bigcup_{i=0}^{\infty} f^{-i}(0)$ of 0 under f, and the set J of accumulation points of I. The set J is a Cantor set and is the Julia set of the polynomial f.

The KNS spectral measure is concentrated on the union of the backward orbits

$$\left(\bigcup_{i=0}^{\infty} f^{-i}(0)\right) \cup \left(\bigcup_{i=0}^{\infty} f^{-i}(-2)\right).$$

The KNS measure of each eigenvalue in $f^{-i}\{0, -2\}$, for $i = 0, 1, \dots$, is $\frac{1}{2 \cdot 3^{i+1}}$.

Remark 3 The Kesten-von-Neumann-Serre measure (KNS measure for short) is the weak limit of the counting spectral measures μ_n associated to the graph Γ_n, for $n \geq 0$ ($\mu_n(B) = m_n(B)/|X|^n$), where $m_n(B)$ counts, including multiplicities, the eigenvalues of Γ_n in B.

For the calculations in this example, the auxiliary pencil of operators used in [23] is 2-dimensional and given by

$$A_n^{(2)}(x, y) = a_n + b_n + c_n + -xI + (y - 1)d_n,$$

where the block structure of d_n is

$$d_n = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}.$$

The recursive formula for the determinant of $A_n^{(2)}(x, y)$ is, for $n \geq 2$,

$$\det(A_n^{(2)}(x, y)) = \left(x^2 - (1 + y)^2\right)^{3^{n-2}} (x^2 - 1 + y - y^2)^{2 \cdot 3^{n-2}} \det(A_{n-1}^{(2)}(F(x, y))),$$

where $F : \mathbb{R}^2 \to \mathbb{R}^2$ is given by

$$F(x, y) = \left(x + \frac{2y^2(-x^2 + x + y^2)}{(x - 1 - y)(x^2 - 1 + y - y^2)}, \frac{y^2(x - 1 + y)}{(x - 1 - y)(x^2 - 1 + y - y^2)}\right).$$

The map $\psi : \mathbb{R}^2 \to \mathbb{R}$ that semi-conjugates F to $f(x) = x^2 - x - 3$ is

$$\psi(x, y) = \frac{x^2 - 1 - xy - 2y^2}{y}.$$

6.3 *The Tangled Odometers Group* $\mathcal{T} = \mathbf{IMG}\left(-\frac{z^3}{2} + \frac{3z}{2}\right)$ *and the First Julia Set*

Theorem 6.3 (Grigorchuk-Nekrashevych-Šunić [16]) *For* $n \geq 0$*, the spectrum of the graph* Γ_n*, as a set, has* $2^{n+1} - 1$ *elements and is equal to*

$$\{4\} \cup \bigcup_{i=0}^{n-1} f^{-i}(2) \cup \bigcup_{j=0}^{n-1} f^{-j}(-2),$$

where

$$f(x) = x^2 - 2x - 4.$$

The multiplicity of the 2^i *eigenvalues in* $f^{-i}(2)$, $i = 0, \ldots, n-1$, *is* 3^{n-1-i}, *the multiplicity of the* 2^j *eigenvalues in* $f^{-j}(-2)$, $j = 0, \ldots, n-1$, *is 1, and the multiplicity of the eigenvalue 4 is 1.*

The spectrum of Γ *(the Schreier spectrum of* $\mathcal{T}$*), as a set, is equal to*

$$\overline{\bigcup_{i=0}^{\infty} f^{-i}(2)}.$$

It consists of a set of isolated points, the backward orbit $I = \bigcup_{i=0}^{\infty} f^{-i}(2)$ *of* 2 *under* f, *and the set* J *of accumulation points of* I. *The set* J *is a Cantor set and is the Julia set of the polynomial* f.

The KNS spectral measure is concentrated on the backward orbit

$$I = \bigcup_{i=0}^{\infty} f^{-i}(2)$$

of f. *The KNS measure of each eigenvalue in* $f^{-i}\{2\}$, *for* $i = 0, 1, \ldots,$ *is* $\frac{1}{3^{i+1}}$.

For the calculations in this example, the auxiliary pencil of operators used in [16] is 3-dimensional and given by

$$A_n^{(3)}(x, y, z) = a_n + b_n + a_n^{-1} + b_n^{-1} - xc_n - (z+2)d_n + (y-1)g_n,$$

where the block structure of c_n, d_n and g_n is

$$c_n = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad d_{n+1} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad e_{n+1} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

6.4 Lamplighter Group $\mathcal{L}_2 = \mathbb{Z} \ltimes \oplus_{\mathbb{Z}} \mathbb{Z}/2\mathbb{Z}$

Theorem 6.4 (Grigorchuk-Żuk [25]) *For* $n \geq 0$, *the spectrum of the graph* Γ_n, *as a set, is equal to*

$$\mathsf{Sp}(\Gamma_n) = \{4\} \cup \left\{ 4\cos\frac{p}{q}\pi \;\mid\; 1 \leq p < q \leq n+1 \text{ and } p \text{ and } q \text{ relatively prime} \right\}.$$

The multiplicity of the eigenvalue $4\cos\frac{p}{q}\pi$, *for* $1 \leq p < q \leq n+1$, *and* p *and* q *relatively prime is equal to*

$$\frac{2^n - 2^{\mathsf{mod}(n,q)}}{2^q - 1} + 1_{\{q \text{ divides } n+1\}},$$

where $\mathsf{mod}(n, q)$ *is the remainder obtained when n is divided by q, and* $1_{\{q \text{ divides } n+1\}}$ *is the indicator function equal to 1 when q divides* $n + 1$ *and to 0 otherwise. The multiplicity of the eigenvalue 4 is 1.*

The spectrum of Γ *(the Schreier spectrum of* $\mathcal{L}_2$*), as a set, is equal to*

$$\mathsf{Sp}(\Gamma) = [-4, 4].$$

The KNS spectral measure is discrete and, for the eigenvalue $4\cos\frac{p}{q}\pi$*, with* $1 \le p < q$ *and p and q relatively prime, is equal to* $\frac{1}{2^q-1}$*.*

The above result has several interesting corollaries. First, note that there exist an infinite ray $\zeta \in \partial X^*$ for which the corresponding parabolic subgroup $P_\zeta = \mathsf{St}_{\mathcal{L}_2}(\zeta)$ is trivial [25] (in fact, this is true for all infinite rays that are not eventually periodic [10, 36]). For such a ray ζ, the Schreier graph $\Gamma_\zeta = \Gamma(\mathcal{L}_2, P_\zeta, S)$ and the Cayley graph $\Gamma(\mathcal{L}_2, S)$ are isomorphic. The calculation of the spectrum of $\mathcal{L}_2$ led to a counterexample of the Strong Atiyah Conjecture. The Strong Atiyah Conjecture states that if $\mathcal{M}$ is a closed Riemannian manifold with fundamental group G, then its L^2-Betti numbers come from the following subgroup of the additive group of rational numbers

$$\mathsf{fin}^{-1}(G) = \left\langle \left\{ \frac{1}{|H|} \mid H \text{ a finite subgroup of } G \right\} \right\rangle \le \mathbb{Q}.$$

This is contradicted by the following result.

Theorem 6.5 (Grigorchuk, Linnell, Schick, Żuk [12]) *There exists a closed Riemannion 7-dimensional manifold* $\mathcal{M}$ *such that all finite groups in its fundamental group G are elementary 2-abelian,* $\mathsf{fin}^{-1}(G) = \mathbb{Z}[\frac{1}{2}]$*, but its third* L^2*-Betti number is* $\beta^3_{(2)}(\mathcal{M}) = \frac{1}{3}$*.*

Note that other versions of Atiyah Conjecture were later also disproved by using examples based on lamplighter-like groups [2, 32].

6.5 *Basilica Group* $\mathcal{B} = \mathsf{IMG}\left(z^2 - 1\right)$ *and* $\mathsf{IMG}\left(z^2 + i\right)$

We do not have complete results for these two examples, but some progress was achieved.

The Schreier spectrum of Basilica group $\mathcal{B}$ was considered in [27], using the auxiliary 2-dimensional pencil of operators given by

$$A_n^{(2)}(x, y) = a_n + a_n^{-1} + y(b_n^{-1} + b_n^{-1}) - xI.$$

Partial results were also obtained by Rogers and Teplyaev by using the spectral decimation method [39].

The group $K = \mathsf{IMG}\left(z^2 + i\right)$ of binary tree automorphisms is generated by three involutions defined by the wreath recursion

$$a = (01)(e, e) \qquad b = ()(a, c) \qquad c = ()(b, e).$$

The Schreier spectrum of $\mathsf{IMG}\left(z^2 + i\right)$ was considered in [24], using the auxiliary 3-dimensional pencil of operators given by

$$A_n^{(3)}(x, y, z) = a_n + yb_n + zc_n - xI.$$

In both cases, the corresponding multi-dimensional map $F : \mathbb{R}^d \to \mathbb{R}^2$ was found, but the shape of the corresponding F-invariant subset (that is, the joint spectrum) is unknown.

7 Laplacians on the Limit Fractals

For some contracting self-similar groups G, the Hecke type operators H_{π_n}, when appropriately rescaled, converge to a well defined *Laplacian* on the limit space. The process of finding the rescaling coefficient and proving existence of the limit Laplacian has much in common with the process of computing the spectra of operators H_{π_n}, as described in Sect. 5. A general theory, working for all contracting groups is still missing, but many interesting examples can be analyzed.

The technique in the known examples is based on the theory of Dirichlet forms on self-similar sets, see [29]. A connection of this theory with self-similar groups, and the examples described in this section are discussed in more detail in [37].

Let G be a self-similar group generated by a finite symmetric set S and, for $n \geq 0$, let $L_n = 1 - H_{\pi_n}$ be the corresponding Laplacian on the Schreier graph Γ_n. Let $\mathcal{E}_n$ be the quadratic form with matrix L_n, that is, the form given by $\mathcal{E}_n(x, y) = \langle L_n x, y\rangle$.

Choose a letter $x_0 \in X$, and consider for every $n \geq 1$ the subset $V_n = x_0^{-\omega} X^n$ of the space $X^{-\omega}$ encoding the limit space of G. We have $V_n \subseteq V_{n+1}$, and we naturally identify V_n with X^n by the bijection $v \mapsto x_0^{-\omega} v$. We also consider $\mathcal{E}_n$ as a form on $\ell^2(V_n) = \ell^2(X^n)$.

The *trace* $\mathcal{E}'_{n+1}$ of $\mathcal{E}_{n+1}$ on V_n is the quadratic form $\mathcal{E}$ such that for $f \in \ell^2(V_n)$ the value of $\mathcal{E}(f, f)$ is equal to the infimum of values of $\mathcal{E}_{n+1}(g, g)$ over all functions $g \in \ell^2(V_{n+1})$ such that $g|_{V_n} = f$.

The matrix of $\mathcal{E}'_{n+1}$ is found as the *Schur complement* of the matrix L_{n+1} of $\mathcal{E}_{n+1}$. Namely, decompose the matrix L_{n+1} into the block form $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$ according to the decomposition of $\ell^2(X^{n+1})$ into the direct sum $\ell^2(x_0 X^n) \oplus \ell^2((X \setminus \{x_0\})X^n)$ (so that A, B, C, and D are of sizes $k^n \times k^n$, $k^n \times (k-1)k^n$, $(k-1)k^n \times k^n$, and $(k-1)k^n \times (k-1)k^n$, respectively, where $k = |X|$ is the size of the alphabet). Then the matrix of $\mathcal{E}'_{n+1}$ is $A - BD^{-1}C$.

Let us consider some examples. Let $G = \mathsf{IMG}\left(z^2 - 1\right)$ be the Basilica group. Consider the Laplacian $1-\alpha(a+a^{-1})-\beta(b+b^{-1})$, and the corresponding Dirichlet forms $\mathcal{E}_n$. Then it follows from the recursive definition of the generators a and b that the decomposition of L_{n+1} into blocks $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$ (for $x_0 = 1$) is

$$\begin{bmatrix} 1-\beta(a+a^{-1}) & -\alpha(1+b^{-1}) \\ -\alpha(1+b) & 1-2\beta \end{bmatrix},$$

hence the matrix of $\mathcal{E}_{n+1,x_0}$ is $\left(1-\frac{\alpha}{2}\right)-\left(\beta(a+a^{-1})+\frac{\alpha}{2}\right)$. Consequently, if we take $\alpha = \frac{2-\sqrt{2}}{2}$, and $\beta = \frac{\sqrt{2}-1}{2}$, then we have $\mathcal{E}'_{n+1} = \lambda\mathcal{E}_n$ for $\lambda = \frac{1}{\sqrt{2}}$. It follows then from the general theory, see [30], that the forms $\lambda^{-n}\mathcal{E}_n$ converge to a Laplacian on the limit space of G, that is, on the Julia set of $z^2 - 1$.

In some cases one needs to take slightly bigger sets V_n. For example, consider the Hanoi Towers group $\mathcal{H}$. Let V_n be the set of sequences of the form $0^{-\infty}X^n$, $1^{-\infty}X^n$, and $2^{-\infty}X^n$. Let $a = (01)(e, e, a)$, $b = (02)(e, b, e)$, and $c = (12)(c, e, e)$, and consider, for positive real numbers x, y, the form $\mathcal{E}_n$ on $\ell^2(V_n)$ given by the matrix

$$\begin{bmatrix} y(1-a)-2x & -x & -x \\ -x & y(1-b)-2x & -x \\ -x & -x & y(1-c)-2x \end{bmatrix}$$

with respect to the decomposition $\ell^2(V_n) = \ell^2(0^{-\omega}X^n) \oplus \ell^2(1^{-\omega}X^n) \oplus \ell^2(2^{-\omega}X^n)$, where a, b, c act on the corresponding subspaces $\ell^2(x^{-\omega}X^n)$ using the representation π_n (after we identify $x^{-\omega}X^n$ with X^n in the natural way).

Then a direct computation using the recursive definition of the generators a, b, c, and the Schur complement shows that trace of $\mathcal{E}_{n+1}$ on V_n is given by the same matrix where (x, y) is replaced by $\left(\frac{3}{5+3x/y}x, y\right)$. Passing to the limit $y \to \infty$, and restricting to functions on which the limit of the quadratic form is finite (which will correspond to identifying sequences $\ldots x^{-\omega}v$ representing the same points of the limit space), we get rescaling $x \mapsto \frac{3}{5}x$, hence convergence of $(5/3)^n\mathcal{E}_n$ to a Laplacian on the limit space of $\mathcal{H}$, which is the Sierpiński gasket.

References

1. G. Amir, O. Angel, B. Virág, Amenability of linear-activity automaton groups. J. Eur. Math. Soc. (JEMS) **15**(3), 705–730 (2013)
2. T. Austin, Rational group ring elements with kernels having irrational dimension. Proc. Lond. Math. Soc. (3) **107**(6), 1424–1448 (2013)
3. L. Bartholdi, R.I. Grigorchuk, On the spectrum of Hecke type operators related to some fractal groups. Tr. Mat. Inst. Steklova **231**(Din. Sist., Avtom. i Beskon. Gruppy), 5–45 (2000)

4. L. Bartholdi, V.A. Kaimanovich, V.V. Nekrashevych, On amenability of automata groups. Duke Math. J. **154**(3), 575–598 (2010)
5. L. Bartholdi, B. Virág, Amenability via random walks. Duke Math. J. **130**(1), 39–56 (2005)
6. I. Bondarenko, R. Grigorchuk, R. Kravchenko, Y. Muntyan, V. Nekrashevych, D. Savchuk, Z. Šunić, On classification of groups generated by 3-state automata over a 2-letter alphabet. Algebra Discrete Math. **7**(1), 1–163 (2008)
7. L. Bartholdi, R. Grigorchuk, V. Nekrashevych, From fractal groups to fractal sets, in *Fractals in Graz 2001*. Trends in Mathematics (Birkhäuser, Basel, 2003), pp. 25–118
8. T. Delzant, R. Grigorchuk, Homomorphic images of branch groups, and Serre's property (FA), in *Geometry and Dynamics of Groups and Spaces*. Progress in Mathematics, vol. 265 (Birkhäuser, Basel, 2008), pp. 351–374
9. M. Fukushima, T. Shima, On a spectral analysis for the Sierpiński gasket. Potential Anal. **1**(1), 1–35 (1992)
10. R. Grigorchuk, R. Kravchenko, On the lattice of subgroups of the lamplighter group. Int. J. Algebra Comput. **24**(6), 837–877 (2014)
11. R. Grigorchuk, D. Lenz, T. Nagnibeda, Spectra of schreier graphs of grigorchuk's group and schroedinger operators with aperiodic order. http://arxiv.org/abs/1412.6822 (2014)
12. R.I. Grigorchuk, P. Linnell, T. Schick, A. Żuk, On a question of Atiyah. C. R. Acad. Sci. Paris Sér. I Math. **331**(9), 663–668 (2000)
13. R. Grigorchuk, V. Nekrashevych, Amenable actions of nonamenable groups. Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) **326**(Teor. Predst. Din. Sist. Komb. i Algoritm. Metody. 13), 85–96, 281 (2005)
14. R. Grigorchuk, V. Nekrashevych. Self-similar groups, operator algebras and Schur complement. J. Modern Dyn. **1**(3), 323–370 (2007)
15. R.I. Grigorchuk, V.V. Nekrashevich, V.I. Sushchanskiĭ, Automata, dynamical systems, and groups. Tr. Mat. Inst. Steklova **231**(Din. Sist., Avtom. i Beskon. Gruppy), 134–214 (2000)
16. R. Grigorchuk, V. Nekrashevych, Z. Šunić, Analysis on the first julia set (2014, in preparation)
17. R.I. Grigorchuk, On Burnside's problem on periodic groups. Funktsional. Anal. i Prilozhen. **14**(1), 53–54 (1980)
18. R.I. Grigorchuk, Degrees of growth of finitely generated groups and the theory of invariant means. Izv. Akad. Nauk SSSR Ser. Mat. **48**(5), 939–985 (1984)
19. R.I. Grigorchuk, Just infinite branch groups, in *New Horizons in Pro-p Groups*, ed. by M.P.F. du Sautoy Dan Segal, A. Shalev (Birkhäuser, Boston, 2000), pp. 121–179
20. N.D. Gupta, S.N. Sidki, On the Burnside problem for periodic groups. Math. Z. **182**(3), 385–388 (1983)
21. R. Grigorchuk, Z. Šuniḱ, Asymptotic aspects of Schreier graphs and Hanoi Towers groups. C. R. Math. Acad. Sci. Paris **342**(8), 545–550 (2006)
22. R. Grigorchuk, Z. Šunić, Self-similarity and branching in group theory, in *Groups St. Andrews 2005. Vol. 1*. London Mathematical Society Lecture Note Series, vol. 339 (Cambridge University Press, Cambridge, 2007), pp. 36–95
23. R. Grigorchuk, Z. Šunić, Schreier spectrum of the Hanoi Towers group on three pegs, in *Analysis on Graphs and Its Applications*. Proceedings of Symposia in Pure Mathematics, vol. 77 (American Mathematical Society, Providence, 2008), pp. 183–198
24. R. Grigorchuk, D. Savchuk, Z. Šunić, The spectral problem, substitutions and iterated monodromy, in *Probability and Mathematical Physics*. CRM Proceedings and Lecture Notes, vol. 42 (American Mathematical Society, Providence, 2007), pp. 225–248
25. R.I. Grigorchuk, A. Żuk, The lamplighter group as a group generated by a 2-state automaton, and its spectrum. Geom. Dedicata **87**(1–3), 209–244 (2001)
26. R.I. Grigorchuk, A. Żuk, On a torsion-free weakly branch group defined by a three state automaton. Int. J. Algebra Comput. **12**(1–2), 223–246 (2002)
27. R.I. Grigorchuk, A. Żuk, Spectral properties of a torsion-free weakly branch group defined by a three state automaton, in *Computational and Statistical Group Theory (Las Vegas, NV/Hoboken, NJ, 2001)*. Contemporary Mathematics, vol. 298 (American Mathematical Society, Providence, 2002), pp. 57–82

28. J.E. Hutchinson, Fractals and self-similarity. Indiana Univ. Math. J. **30**(5), 713–747 (1981)
29. J. Kigami, *Analysis on Fractals*. Cambridge Tracts in Mathematics, vol. 143 (Cambridge University Press, Cambridge/New York, 2001)
30. J. Kigami, Harmonic analysis for resistance forms. J. Funct. Anal. **204**(2), 399–444 (2003)
31. T. Kumagai, Regularity, closedness and spectral dimensions of the Dirichlet forms on P.C.F. self-similar sets. J. Math. Kyoto Univ. **33**(3), 765–786 (1993)
32. F. Lehner, S. Wagner, Free lamplighter groups and a question of Atiyah. Am. J. Math. **135**(3), 835–849 (2013)
33. L. Malozemov, Spectral theory of the differential Laplacian on the modified Koch curve, in *Geometry of the Spectrum (Seattle, WA, 1993)*. Contemporary Mathematics, vol. 173 (American Mathematical Society, Providence, 1994), pp. 193–224
34. L. Malozemov, A. Teplyaev, Self-similarity, operators and dynamics. Math. Phys. Anal. Geom. **6**(3), 201–218 (2003)
35. V. Nekrashevych, *Self-Similar Groups*. Mathematical Surveys and Monographs, vol. 117 (American Mathematical Society, Providence, 2005)
36. V. Nekrashevych, G. Pete, Scale-invariant groups. Groups Geom. Dyn. **5**(1), 139–167 (2011)
37. V. Nekrashevych, A. Teplyaev, Groups and analysis on fractals, in *Analysis on Graphs and its Applications*, ed. by P. Exner, J.P. Keating, P. Kuchment, T. Sunada, A. Teplyaev. Proceedings of Symposia in Pure Mathematics, vol. 77 (American Mathematical Society, Providence, 2008), pp. 143–170
38. R. Rammal, G. Toulouse, Spectrum of the Schrödinger equation on a self-similar structure. Phys. Rev. Lett. **49**(16), 1194–1197 (1982)
39. L.G. Rogers, A. Teplyaev, Laplacians on the basilica Julia sets. Commun. Pure Appl. Anal. **9**(1), 211–231 (2010)
40. T. Shima, On eigenvalue problems for Laplacians on p.c.f. self-similar sets. Jpn. J. Indust. Appl. Math. **13**(1), 1–23 (1996)
41. S. Sidki, Automorphisms of one-rooted trees: growth, circuit structure, and acyclicity. J. Math. Sci. (New York) **100**(1), 1925–1943 (2000)
42. R.S. Strichartz, *Differential Equations on Fractals* (Princeton University Press, Princeton, 2006). A tutorial
43. A. Teplyaev, Spectral analysis on infinite Sierpiński gaskets. J. Funct. Anal. **159**(2), 537–567 (1998)
44. A. Teplyaev, Spectral zeta function of symmetric fractals, in *Fractal Geometry and Stochastics III*. Progress in Probability, vol. 57 (Birkhäuser, Basel, 2004), pp. 245–262

Finite Energy Coordinates and Vector Analysis on Fractals

Michael Hinz and Alexander Teplyaev

Abstract We consider local finite energy coordinates associated with a strongly local regular Dirichlet form on a metric measure space. We give coordinate formulas for substitutes of tangent spaces, for gradient and divergence operators and for the infinitesimal generator. As examples we discuss Euclidean spaces, Riemannian local charts, domains on the Heisenberg group and the measurable Riemannian geometry on the Sierpinski gasket.

Keywords Metric space • Dirichlet form • Coordinates • Differential forms

Mathematics Subject Classification (2000). Primary 28A80, 31E05, 47A07, 60J25, 60J35; Secondary 35P20, 47A75, 81Q10, 81Q35

1 Introduction

Suitable coordinate maps are tools in many branches of geometry. For instance, smooth coordinate changes are the crucial ingredient in the definition of a differentiable structure on a manifold and therefore omnipresent in differential geometry (e.g. [26, 29]). For a general metric measure space we can not expect to find local coordinates that transform smoothly. However, in the field of analysis on fractals Kusuoka [35, 36], Kigami [32, 34], Strichartz [42], Teplyaev [44], Hino [15, 16],

---

Research supported in part by NSF grant DMS-0505622 and by SFB 701 of the German Research Council (DFG).

M. Hinz (✉)
Department of Mathematics, Bielefeld University, P.O. Box 100131, D-33501 Bielefeld, Germany
e-mail: mhinz@math.uni-bielefeld.de

A. Teplyaev
Department of Mathematics, University of Connecticut, Storrs, CT 06269-3009, USA
e-mail: Alexander.Teplyaev@uconn.edu

C. Bandt et al. (eds.), *Fractal Geometry and Stochastics V*, Progress in Probability 70, DOI 10.1007/978-3-319-18660-3_12

Kajino [30, 31] and others have contributed to a concept that is now referred to as 'measurable Riemannian geometry'. This concept is based on Dirichlet forms and involves the use of harmonic functions as 'global coordinates'. In probability similar ideas can already be found in works of Doob, Dynkin and Skorohod. On the other hand there is recent progress in the studies of a first order calculus on fractals, [8, 9, 19–22, 25, 28] again based on Dirichlet form theory, [6, 13], which allows to discuss differential 1-forms and vector fields, partially based on [7, 40].

In the present note we consider metric measure spaces, equipped with a strongly local Dirichlet form and consider associated local finite energy coordinates. Analogously to Riemannian geometry, we provide coordinate expressions for the gradient and divergence operators (derivation and coderivation) used in the first order theory, and for the infinitesimal generator. The present paper is a brief introduction to the subject. We hope to facilitate understanding of how the measurable first order calculus is related to Euclidean, Riemannian, sub-Riemannian and measurable Riemannian situations.

2 Preliminaries

Let X be a locally compact separable metric space and μ a nonnegative Radon measure on X such that $\mu(U) > 0$ for all nonempty open $U \subset X$. Let $(\mathcal{E}, \mathcal{F})$ be a *strongly local regular Dirichlet form* on $L_2(X, \mu)$, that is:

(1) $\mathcal{F}$ is a dense subspace of $L_2(X, \mu)$ and $\mathcal{E} : \mathcal{F} \times \mathcal{F} \to \mathbb{R}$ is a nonnegative definite symmetric bilinear form; we denote $\mathcal{E}(f) := \mathcal{E}(f,f)$;
(2) $\mathcal{F}$ is a Hilbert space with the norm $\sqrt{\mathcal{E}_1(f)} := \left(\mathcal{E}(f) + \|f\|^2_{L_2(X,\mu)}\right)^{1/2}$;
(3) For any $f \in \mathcal{F}$ we have $(f \vee 0) \wedge 1 \in \mathcal{F}$ and $\mathcal{E}((f \vee 0) \wedge 1) \le \mathcal{E}(f)$, where $f \vee g := \max\{f, g\}$ and $f \wedge g := \min\{f, g\}$;
(4) $\mathcal{C} := \mathcal{F} \cap C_c(X)$ is dense both in $\mathcal{F}$ with respect to the norm $\sqrt{\mathcal{E}_1(f)}$, and in the space $C_c(X)$ of continuous compactly supported functions with respect to the uniform norm;
(5) If $f, g \in \mathcal{C}$ and g is constant on a neighborhood of supp$f$ then $\mathcal{E}(f, g) = 0$.

To each Dirichlet form $(\mathcal{E}, \mathcal{F})$ on $L_2(X, \mu)$ there exists a unique non-positive self-adjoint operator $(L, \mathrm{dom}\, L)$, called the *infinitesimal generator of* $(\mathcal{E}, \mathcal{F})$, such that $\mathrm{dom}\ L \subset \mathcal{F}$ and

$$\mathcal{E}(f, g) = -\langle Lf, g\rangle_{L_2(X,\mu)}$$

for all $f \in \mathrm{dom}\ L$ and $g \in \mathcal{F}$. See [6, 13]. For pointwise products we have

$$\mathcal{E}(fg)^{1/2} \le \mathcal{E}(f)^{1/2}\, \|g\|_{L_\infty(X,\mu)} + \mathcal{E}(g)^{1/2}\, \|f\|_{L_\infty(X,\mu)}\,, \quad f, g \in \mathcal{F} \cap L_\infty(X, \mu), \tag{2.1}$$

[5, Corollary I.3.3.2], and in particular, the space $\mathcal{C} := \mathcal{F} \cap C_c(X)$ is an algebra. For any $f, g \in \mathcal{C}$ a signed Radon measure $\Gamma(f, g)$ on X is defined by

$$\int_X \varphi d\Gamma(f, g) = \frac{1}{2}\big(\mathcal{E}(f, g\varphi) + \mathcal{E}(g, f\varphi) - \mathcal{E}(fg, \varphi)\big), \quad \varphi \in \mathcal{C}. \tag{2.2}$$

By approximation in $\mathcal{F}$ we also define $\Gamma(f, g)$ for any $f, g \in \mathcal{F}$, referred to as the *(mutual) energy measure of f and g*, see [13]. We denote the nonnegative measure $\Gamma(f) = \Gamma(f, f)$. Below it will be advantageous to consider functions that are only locally of finite energy. We define $\mathcal{F}_{loc}$ as the set of functions $f \in L_{2,loc}(X, \mu)$ such that for any relatively compact open set $V \subset X$ there exists some $u \in \mathcal{F}$ such that $f|_V = u|_V$ μ-a.e. Exhausting X by an increasing sequence of relatively compact open sets and using related cut-off functions we can define $\Gamma(f)$ for $f \in \mathcal{F}_{loc}$. If V is relatively compact open and $u \in \mathcal{F}$ agrees with f μ-a.e. on V then

$$\Gamma(f)|_V = \Gamma(u)|_V. \tag{2.3}$$

Example 1 A prototype for a strongly local regular Dirichlet form is the *Dirichlet integral*

$$\mathcal{E}(f) = \int_{\mathbb{R}^n} |\nabla f|^2 \, dx$$

on $L_2(\mathbb{R}^n)$, where $\mathcal{F}$ is the Sobolev space $H^1(\mathbb{R}^n)$ of functions $f \in L_2(\mathbb{R}^n)$ with $\frac{\partial f}{\partial x_i} \in L_2(\mathbb{R}^n)$ for all i. Note that $C_c^1(\mathbb{R}^n)$ is dense in $H^1(\mathbb{R}^n)$ and in $C_c(\mathbb{R}^n)$. The generator is the Laplacian $L = \Delta$ and the energy measures are given by $\Gamma(f) = |\nabla f|^2 \, dx$.

A nonnegative Radon measure m on X is called *energy dominant* if all energy measures $\Gamma(f), f \in \mathcal{F}$, are absolutely continuous with respect to m, [15, 16, 18, 25]. By $\frac{d\Gamma(f)}{dm}$ we denote the corresponding Radon-Nikodym densities.

Let $\varphi \in \mathcal{C}$, let V be a relatively compact open neighborhood V of $\operatorname{supp}\varphi$ and suppose $(f_n)_n \subset \mathcal{F}_{loc}$. We say that φ *is locally approximated by the sequence* $(f_n)_n$ *on* V if there is a sequence $(u_n)_n \subset \mathcal{F}$ with $\lim_n \mathcal{E}_1(\varphi - u_n) = 0$ and $f_n|_V = u_n|_V$ μ-a.e. for all n. The following lemma follows from (2.3), [13, Theorem 2.1.4 and Lemma 3.2.4].

Lemma 2.1 *Let $\varphi \in \mathcal{C}$ and let V be a relatively compact open neighborhood V of* $\operatorname{supp}\varphi$. *Suppose that φ is locally approximated by $(f_n)_n \subset \mathcal{F}_{loc}$ on V. Then* $\lim_n \Gamma(\varphi - f_n)(V) = 0$. *If the functions f_n are continuous on V, then there is a subsequence $(f_{n_k})_k$ such that* $\lim_k f_{n_k} = \varphi$ *$\Gamma(f)$-a.e. on V for any $f \in \mathcal{F}_{loc}$.*

3 Finite Energy Coordinates

Let $y = (y^i)_{i\in I}$ be a finite or countable collection of locally bounded functions y^i. Given a finite ordered subset $J = (n_1, \ldots, n_k)$ of I, the space of all functions of form $f = F(y^{n_1}, \ldots, y^{n_k})$, where the functions F are polynomials in k variables and such that $F(0) = 0$, will be denoted by $\mathcal{P}_J(y)$. For a fixed collection $(y^i)_{i\in I}$ set

$$\mathcal{P}(y) := \bigcup_{J\subset I} \mathcal{P}_J(y), \tag{3.1}$$

the union taken over all ordered finite subsets J of I. Note that $\mathcal{P}(y)$ is an algebra of locally bounded functions. For any k we regard the space $\mathbb{R}^k$ as a subspace of $\mathbb{R}^{\mathbb{N}}$ containing $(v_1, v_2, \ldots, v_k, 0, 0, \ldots)$ for $(v_1, v_2, \ldots, v_k) \in \mathbb{R}^k$. Similarly, we consider $(k \times k)$-matrices as linear operators from $\mathbb{R}^{\mathbb{N}}$ to $\mathbb{R}^{\mathbb{N}}$.

Definition 3.1 Let m be an energy dominant measure for $(\mathcal{E}, \mathcal{F})$. A finite or countable collection $y = (y^i)_{i\in I}$ of continuous and locally bounded functions $y^i \in \mathcal{F}_{loc}$ is called a *coordinate sequence for* $(\mathcal{E}, \mathcal{F})$ *with respect to m* if

(i) Any $\varphi \in \mathcal{C}$ can locally be approximated on a relatively compact neighborhood V of $\operatorname{supp}\varphi$ by a sequence of elements of $\mathcal{P}(y)$

(ii) For any $i \in I$ we have

$$\frac{d\Gamma(y^i)}{dm} \in L_1(X, m) \cap L_\infty(X, m),$$

and for any i and j

$$Z^{ij}(x) := \frac{d\Gamma(y^i, y^j)}{dm}(x)$$

are Borel functions (versions) such that for m-a.e. $x \in X$, $Z(x) := (Z^{ij}(x))_{ij=1}^{\infty}$ defines a bounded symmetric nonnegative definite linear operator $Z(x) : l_2 \to l_2$.

(iii) We say that *the coordinates* y^i *have finite energy* if $y^i \in \mathcal{F}$ for all $i \in I$.

A coordinate sequence $y = (y^i)_{i\in I}$ induces a mapping $y : X \to \mathbb{R}^{\mathbb{N}}$.

Remark 1 Condition (i) in Definition 3.1 makes sense because we have $\mathcal{P}(y) \subset \mathcal{F}_{loc}$. If the coordinates y^i have finite energy the inclusion $\mathcal{P}(y) \subset \mathcal{F}$ is clear from (2.1). To see this inclusion in the general case it suffices to show that for any continuous and locally bounded $f, g \in \mathcal{F}_{loc}$ we have $fg \in \mathcal{F}_{loc}$. Clearly $fg \in L_{2,loc}(X, \mu)$. Further, given a relatively compact open set $V \subset X$ we can find a suitable cutoff function $\chi \in \mathcal{C}$ with $0 \le \chi \le 1$ and $\chi \equiv 1$ on V, a relatively compact open neighborhood of $\operatorname{supp}\chi$ and functions $u, v \in \mathcal{F}$ such that $f|_U = u|_U$ and

$g|_U = v|_U$ μ-a.e. Clearly $\chi u \in L_2(X,\mu)$, and using locality, [13, Corollary 3.2.1],

$$\mathcal{E}(\chi u)^{1/2} = \left(\int_U d\Gamma(\chi u)\right)^{1/2} \leq \left(\int_U \chi^2 d\Gamma(u)\right)^{1/2} + \left(\int_U \tilde{u}^2 d\Gamma(\chi)\right)^{1/2},$$

where $\tilde{u}$ is a quasi-continuous version of u. See e.g. [13, Chapter II] for quasi-continuity and the Appendix in [24] for comments on the formula (which also follows from Cauchy-Schwarz applied to (4.5) below). Approximating u in $\mathcal{E}_1^{1/2}$-norm by a sequence from $\mathcal{C}$ we see that χu is the limit in $\mathcal{E}_1^{1/2}$-norm of a sequence from $\mathcal{C}$, and by completeness χu is in $\mathcal{F}$. Similarly for χv. Both functions are bounded on U μ-a.e. and vanish outside U, hence are also members of $L_2(X,\mu)$. Therefore $\chi^2 uv \in \mathcal{F}$ by (2.1), what implies $fg \in \mathcal{F}_{loc}$.

In Sect. 8 we show that (under an additional continuity assumption) it is always possible to construct a finite energy dominant measure and a corresponding coordinate sequence of finite energy coordinates. The following examples relate Definition 3.1 to well known situations.

Example 2

(1) Consider

$$\mathcal{E}(f) := \sum_{i,j=1}^{n} \int_{\mathbb{R}^n} a_{ij}(x) \frac{\partial f}{\partial x_i}(x) \frac{\partial f}{\partial x_j}(x) dx, \ \ f \in C_c^1(\mathbb{R}^n),$$

where $a_{ij} = a_{ji}$ are bounded Borel functions satisfying $\sum_{i,j=1}^n a_{ij}(x)\xi_i\xi_j \geq c|\xi|^2$ with a universal constant $c > 0$ for any $\xi \in \mathbb{R}^n$ and λ^n-a.e. $x \in \mathbb{R}^n$. Here λ^n denotes the n-dimensional Lebesgue measure $\lambda^n(dx) = dx$. Then $(\mathcal{E}, C_c^1(\mathbb{R}^n))$ is closable in the space $L_2(\mathbb{R}^n)$, and its closure $(\mathcal{E}, H^1(\mathbb{R}^n))$ is a strongly local regular Dirichlet form. Obviously λ^n is energy dominant for $(\mathcal{E}, H^1(\mathbb{R}^n))$. The Euclidean coordinates $y^k(x) = x_k$, $k = 1, \ldots, n$, form a coordinate sequence for $(\mathcal{E}, H^1(\mathbb{R}^n))$ with respect to λ^n. Note that $\nabla y^k = e_k$ is the k-th unit vector in $\mathbb{R}^n$, and we have

$$Z^{ij}(x) = a_{ij}(x) \quad \text{for } \lambda^n\text{-a.e. } x \in \mathbb{R}^n$$

and $i,j = 1,\ldots n$. This shows (ii). If $\varphi \in C_c^1(\mathbb{R}^n)$ then we can find a relatively compact open set V containing $\operatorname{supp}\varphi$ on which the function φ can be approximated it in C^1-norm by a sequence of polynomials in the variables $x_1, \ldots, x_n$, hence in the coordinates $y^1, \ldots, y^n$. Multiplying these polynomials by a (nonnegative) C^1-cut-off function supported in V and equal to one on $\operatorname{supp}\varphi$, the approximation is seen to take place in $H^1(\mathbb{R}^n)$. As $C_c^1(\mathbb{R}^n)$ is dense in $H^1(\mathbb{R}^n)$, this implies (i). The coordinates y^k do not have finite energy.

(2) Let (M,g) be an n-dimensional Riemannian manifold, [26, 29], let (V,y) be a local chart with coordinates $y = (y^1, \ldots, y^n)$ and U a relatively compact open

set with $\overline{U} \subset V$. By $dvol$ we denote the Riemannian volume (restricted to U). The closure $(\mathcal{E}, \mathring{H}^1(U))$ in $L_2(U, dvol)$ of

$$\mathcal{E}(f) := \int_U \langle \nabla f, \nabla f \rangle_{T_x M} \, dvol(x), \ \ f \in C_c^1(U).$$

is a strongly local Dirichlet form. The reference measure $dvol$ is energy dominant, for any $k = 1, \ldots, n$ we have

$$\nabla y^k = g^{kj} \frac{\partial}{\partial y^j}$$

and

$$Z^{kk} = \left\langle \nabla y^k, \nabla y^k \right\rangle_{TM} = g^{kj} g^{ki} \left\langle \frac{\partial}{\partial y^i}, \frac{\partial}{\partial y^j} \right\rangle_{TM} = g^{kk}$$

and therefore (ii). Recall that

$$\nabla f = g^{ij} \frac{\partial f}{\partial y^i} \frac{\partial}{\partial y^j}, \ \ g_{ij} = \left\langle \frac{\partial}{\partial y^i}, \frac{\partial}{\partial y^j} \right\rangle_{TM}$$

and $g^{ki} g_{ij} = \delta_j^k$. For a function $f \in C_c^1(U)$ the function $f \circ y^{-1}$ is a member of $C^1(W)$, and accordingly it can be approximated in $C^1(W)$-norm by a sequence $(p_m)_m$ of polynomials in the variables $y^1, \ldots, y^n$. Consequently the functions $p_m \circ y$ approximate f in $C^1(U)$-norm (note that the differentials $d((p_m \circ y) \circ y^{-1})(y(x))$ approximate $d(f \circ y^{-1})(y(x))$ uniformly in $x \in U$). This implies (i). Here the y^i are not in $\mathcal{F}$ because they do not satisfy the Dirichlet boundary conditions on ∂U.

(3) A sub-Riemannian example is given by the Heisenberg group $\mathbb{H}$, [10, 27, 38, 41], realized as $\mathbb{R}^3$ together with the non-commutative multiplication

$$(\xi_1, \eta_1, \zeta_1) \cdot (\xi_2, \eta_2, \zeta_2) := (\xi_1 + \xi_2, \eta_1 + \eta_2, \zeta_1 + \zeta_2 + \xi_1 \eta_2 - \eta_1 \xi_2).$$

Left multiplication by $(\xi, 0, 0)$ and $(0, \eta, 0)$ yields the left-invariant vector fields

$$X(q) := \frac{\partial}{\partial \xi}\Big|_q - \frac{1}{2} \eta \frac{\partial}{\partial \zeta}\Big|_q \quad \text{and} \quad Y(q) := \frac{\partial}{\partial \eta}\Big|_q + \frac{1}{2} \xi \frac{\partial}{\partial \zeta}\Big|_q,$$

and at each $q = (\xi, \eta, \zeta) \in \mathbb{H}$ the tangent vectors $X(q)$ and $Y(q)$ span a two-dimensional subspace V_q of the tangent space $T_q \mathbb{H} \cong \mathbb{R}^3$. The sub-Riemannian metric is given by the inner products $\langle \cdot, \cdot \rangle_{V_q}$ on the spaces V_q that makes $(X(q), Y(q))$ an orthonormal basis, respectively. We use the Haar measure on $\mathbb{H}$, which coincides with the Lebesgue measure λ^3 on $\mathbb{R}^3$. Now let $U \subset \mathbb{H}$ be a connected bounded open set and consider the bilinear form

$$\mathcal{E}(f) := \int_U ((Xf)^2 + (Yf)^2)\, d\lambda^3,\ \ f \in C_c^1(U).$$

Let $(\mathcal{E}, \mathring{S}^1(U))$ denote the closure of $(\mathcal{E}, C_c^1(U))$ in $L_2(U)$. Obviously λ^3 is energy dominant. A coordinate sequence for $(\mathcal{E}, \mathring{S}^1(U))$ and λ^3 is given by $y = (y^1, y^2, y^3) := (\xi, \eta, \zeta)$. Condition (i) follows again by polynomial approximation in $C_c^1(U)$. It is immediate that $Xy^1 = 1$, $Yy^1 = 0$, similarly for y^2, and $Xy^3 = -\frac{\eta}{2}$, $Yy^3 = \frac{\xi}{2}$, which yields the symmetric and nonnegative definite matrices

$$Z(q) = \begin{pmatrix} 1 & 0 & -\frac{\eta}{2} \\ 0 & 1 & \frac{\xi}{2} \\ -\frac{\eta}{2} & \frac{\xi}{2} & \frac{\xi^2+\eta^2}{4} \end{pmatrix},$$

so that (ii) is satisfied. For any $q \in \mathbb{H}$ the matrix $Z(q)$ has rank 2. As in (2) the coordinates are not in $\mathcal{F}$.

(4) We consider a prototype of a finitely ramified fractal in finite energy coordinates. Let K denote the Sierpinski gasket, seen as the post-critically self-similar structure generated by the maps $f_j : \mathbb{R}^2 \to \mathbb{R}^2, f_j(x) = \frac{1}{2}(x + p_j), j = 1, 2, 3,$ where p_1, p_2 and p_3 are the vertices of an equilateral triangle in $\mathbb{R}^2$. Let $(\mathcal{E}, \mathcal{F})$ be the standard resistance form on K, obtained as the rescaled limit of discrete energy forms along a sequence of graphs with increasing vertex sets V_n 'approximating K',

$$\mathcal{E}(f) = \lim_{n\to\infty} \left(\frac{5}{3}\right)^n \sum_{p,q\in V_n} (f(p) - f(q))^2,$$

see e.g. [32, 33, 35, 36, 43] for details. With $\{p_1, p_2, p_3\}$ as boundary and with Dirichlet boundary conditions there exist two harmonic functions $y^1, y^2 \in \mathcal{F}$ with $\mathcal{E}(y^1) = \mathcal{E}(y^2) = 1$ and $\mathcal{E}(y^1, y^2) = 0$ such that the mapping $y : K \to \mathbb{R}^2$

$$y(x) := (y^1(x), y^2(x)),\ \ x \in K, \tag{3.2}$$

is a homeomorphism from K onto its image $y(K) \subset \mathbb{R}^2$. We consider K endowed with the *Kusuoka measure* ν, defined as the sum

$$\nu := \Gamma(y^1) + \Gamma(y^2)$$

of the energy measure $\Gamma(y^1)$ and $\Gamma(y^2)$ of y^1 and y^2, respectively. The resistance form $(\mathcal{E}, \mathcal{F})$ induces a strongly local Dirichlet form on $L_2(K, \nu)$, for which the finite measure ν is energy dominant. The pair (y^1, y^2) is a coordinate sequence for this form: Condition (ii) is satisfied by construction, condition (i) follows by polynomial approximation and the density of functions of type $F \circ y$, $F \in C^1(\mathbb{R}^2)$, in $\mathcal{F}$, see e.g. [32, 35, 36, 45]. The operators $Z(x)$ may be viewed as (2×2)-matrices, and for ν-a.e. $x \in K$ the matrix $Z(x)$ is symmetric, nonnegative definite and has rank 1.

4 Energy, Fibers and Bundles

In what follows we will assume throughout that $(\mathcal{E}, \mathcal{F})$ is a strongly local regular Dirichlet form on $L_2(X, \mu)$, m is an energy dominant measure and $y = (y^i)_{i\in I}$ be a coordinate sequence for $(\mathcal{E}, \mathcal{F})$ with respect to m.

We would like to emphasize that unless stated otherwise we do not assume that the reference measure itself is energy dominant or that the form $(\mathcal{E}, \mathcal{F})$ has a restriction that is closable with respect to the energy dominant measure m under consideration.

In Example 2 (4), a well known formula of Kusuoka [35] and Kigami [32] is

$$\mathcal{E}(f, g) = \int_K \langle \nabla F(y), Z(x)\nabla G(y)\rangle_{\mathbb{R}^2}\, \nu(dx), \tag{4.1}$$

for all $f = F \circ y$ and $g = G \circ y$ with $F, G \in C^1(\mathbb{R}^2)$. This identity expresses the *energy* in terms of coordinates. As the matrix Z varies measurably in x, it has been named a *measurable Riemannian metric*, [15, 30, 34]. The following is version of (4.1) immediately following from the chain rule [13, Theorem 3.3.2].

Lemma 4.1 *Let m be an energy dominant measure and $(y^i)_{i\in I}$ a coordinate sequence. For all $f = F \circ y$ and $g = G \circ y$ from $\mathcal{P}(y)$ we have*

$$\Gamma(f, g)(x) = \langle \nabla F(y), Z(x)\nabla G(y)\rangle_{l_2} \tag{4.2}$$

for m-a.e. $x \in X$. If in addition $f, g \in \mathcal{F}$, then

$$\mathcal{E}(f, g) = \int_X \langle \nabla F, Z\nabla G\rangle_{l_2}\, dm.$$

We rewrite (4.2) in a somewhat artificial way. For any $x \in X$ such that $Z(x)$ is symmetric and nonnegative definite, the bilinear extension of

$$\langle f_1 \otimes g_1, f_2 \otimes g_2\rangle_{\mathcal{H}_x} := G_1(y)G_2(y)\, \langle \nabla F_1(y), Z(x)\nabla F_2(y)\rangle_{l_2}\,, \tag{4.3}$$

where $f_i = F_i \circ y$ and $g_i = G_i \circ y$ are members of $\mathcal{P}(y)$ with polynomials F_i and G_i, $i = 1, 2$, defines a nonnegative definite symmetric bilinear form on the vector space $\mathcal{P}(y) \otimes \mathcal{P}(y)$. Let $\|\cdot\|_{\mathcal{H}_x}$ denote the associated Hilbert seminorm. Factoring out zero seminorm elements and completing, we obtain a Hilbert space $(\mathcal{H}_x, \langle\cdot,\cdot\rangle_{\mathcal{H},x})$. The

$\mathcal{H}_x$-equivalence class of an element $f \otimes g$ of $\mathcal{P}(y) \otimes \mathcal{P}(y)$ we denote by $(f \otimes g)_x$. Note that for m-a.e. $x \in X$ the expression in (4.3) equals

$$g_1(x)g_2(x)\frac{\Gamma(f_1,f_2)}{dm}(x).$$

Example 3

(1) In the situation of Example 2 (1) we observe $\mathcal{H}_x \cong \mathbb{R}^n$ for λ^n-a.e. $x \in \mathbb{R}^n$ and

$$\langle f_1 \otimes g_1, f_2 \otimes g_2\rangle_{\mathcal{H}_x} = g_1(x)g_2(x)\,\langle \nabla f_1(x), a(x)\nabla f_2(x)\rangle_{\mathbb{R}^n}\,,$$

where we write $a = (a_{ij})_{i,j=1}^n$.

(2) For the Riemannian situation in Example 2 (2) we have

$$\langle f_1 \otimes g_1, f_2 \otimes g_2\rangle_{\mathcal{H}_x} = g_1(x)g_2(x)\,\langle df_1(x), df_2(x)\rangle_{T_x^*M}$$

for $dvol$-a.e. $x \in U$, where

$$f \mapsto df = \sum_{i=1} \frac{\partial f}{\partial y^i} dy^i \tag{4.4}$$

denotes the exterior derivation. Note that $\mathcal{H}_x \cong T_x^*M \cong T_xM \cong \mathbb{R}^n$.

(3) For the Heisenberg group as in Example 2 (3),

$$\langle f_1 \otimes g_1, f_2 \otimes g_2\rangle_{\mathcal{H}_q} = g_1(q)g_2(q)\,((X(q)f_1)(X(q)f_2) + (Y(q)f_1)(Y(q)f_2))$$

for λ^3-a.e. $q \in U$. Here $\mathcal{H}_q$ is isometrically isomorphic to the horizontal fiber V_q.

We proceed to a more global perspective. A nonnegative definite symmetric bilinear form on $\mathcal{C} \otimes \mathcal{C}$ can be introduced by extending

$$\langle f_1 \otimes g_1, f_2 \otimes g_2\rangle_{\mathcal{H}} := \int_X g_1(x)g_2(x)\,\Gamma(f_1,f_2)(x)m(dx). \tag{4.5}$$

The associated Hilbert seminorm is denoted by $\|\cdot\|_{\mathcal{H}}$. Factoring out zero seminorm elements and completing yields another Hilbert space $\mathcal{H}$, usually referred to a the *Hilbert space of* 1*-forms associated with* $(\mathcal{E}, \mathcal{F})$. This definition has some history, see e.g. [11, 39, 40], and in the context of Dirichlet forms it was first introduced by Cipriani and Sauvageot in [7]. Right and left actions of $\mathcal{C}$ on the space $\mathcal{C} \otimes \mathcal{C}$ can be defined by extending

$$(f \otimes g)h := f \otimes (gh) \quad \text{and} \quad h(f \otimes g) = (fh) \otimes g - h \otimes (fg). \tag{4.6}$$

By strong locality they coincide. Moreover, they extend further to an action of $\mathcal{C}$ on $\mathcal{H}$ and $\|\omega h\|_{\mathcal{H}} \leq \|h\|_{L_\infty(X,m)} \|\omega\|_{\mathcal{H}}$ for any $\omega \in \mathcal{H}$ and $h \in \mathcal{C}$. A linear operator $\partial : \mathcal{C} \to \mathcal{H}$ can be introduced by setting

$$\partial f := f \otimes \mathbf{1}, \ \ f \in \mathcal{C},$$

note that $f \otimes \mathbf{1}$ is a member of $\mathcal{H}$, as can be seen from (4.5) by approximating $\mathbf{1}$ pointwise. The operator ∂ is a derivation, i.e.

$$\partial(fg) = (\partial f)g + f\partial g, \ \ f, g \in \mathcal{C}. \tag{4.7}$$

It satisfies

$$\|\partial f\|^2_{\mathcal{H}} = \mathcal{E}(f), \ \ f \in \mathcal{C}, \tag{4.8}$$

and extends to a closed unbounded operator $\partial : L_2(X, \mu) \to \mathcal{H}$ with domain $\mathcal{F}$.

Since the left action in (4.6) is also well defined for bounded Borel functions, approximation shows that $(f \otimes g)\mathbf{1}_V = (\partial f)g\mathbf{1}_V$ is in $\mathcal{H}$ for any $f, g \in \mathcal{F}$ and relatively compact open V. By locality, (2.3) and approximation (pointwise m-a.e.) we then have $(f \otimes g)\mathbf{1}_V \in \mathcal{H}$ even for locally bounded $f, g \in \mathcal{F}_{loc}$. Formulas (4.6) and (4.7) have local versions valid for elements of $\mathcal{P}(y)$. Note also that for m-a.e $x \in X$,

$$\langle (\partial f)_x, (\partial g)_x \rangle_{\mathcal{H}_x} = \frac{d\Gamma(f, g)}{dm}(x).$$

Let $\{\kappa_N\}_{N=1}^\infty$ be a family of bounded Lipschitz functions $\kappa_N : \mathbb{R} \to \mathbb{R}$ such that for all N we have $\kappa_N(t) = t$ on $[-N, N]$. Let $\tilde{\mathcal{P}}(y)$ denote the collection of functions $\kappa_N \circ g$ for any $g \in \mathcal{P}(y)$ and N. The next lemma contains a version of Lemma 4.1.

Lemma 4.2 *For $f_i = F_i \circ y$ and $g_i = G_i \circ y$ from $\mathcal{P}(y)$, $i = 1, 2$, and any relatively compact open V we have*

$$\begin{aligned} \langle (f_1 \otimes g_1)\mathbf{1}_V, f_2 \otimes g_2 \rangle_{\mathcal{H}} &= \int_V \langle (f_1 \otimes g_1)_x, (f_2 \otimes g_2)_x \rangle_{\mathcal{H}_x} m(dx) \\ &= \int_V G_1(y)G_2(y) \langle \nabla F_1(y), Z(x)\nabla F_2(y) \rangle_{l_2} m(dx). \end{aligned}$$

If in addition $f, g \in \mathcal{F}$, then we can replace V by X. Moreover,

$$\operatorname{span}\left(\left\{(f \otimes g)\mathbf{1}_V : f \in \mathcal{P}(y), \ g \in \tilde{\mathcal{P}}(y), \ V \subset X \textit{ relatively compact open}\right\}\right)$$

is a dense subspace of $\mathcal{H}$. If the coordinates y^i have finite energy, then $\mathcal{P}(y) \otimes \tilde{\mathcal{P}}(y)$ is a dense subspace of $\mathcal{H}$.

Proof The first statement is obvious. To see the second, let φ and ψ be functions from $\mathcal{C}$ and U a relatively compact open set containing $\operatorname{supp}\varphi$ on which φ is locally approximated on U by a sequence $(f_n)_n \subset \mathcal{P}(y)$. We have $\|\varphi \otimes \psi - f_n \otimes \psi\|_{\mathcal{H}}^2 \leq \sup_{x\in X} |\psi(x)|^2 \Gamma(\varphi - f_n)(U)$, which converges to zero by Lemma 2.1. Hence the span of elements $f \otimes \psi$ with $f \in \mathcal{P}(y)$ and $\psi \in \mathcal{C}$ is dense in $\mathcal{H}$. On the other hand, if V is a relatively compact open set containing $\operatorname{supp}\psi$ and $(g_n)_n \subset \mathcal{P}(y)$ approximates ψ locally on V, after replacing the sequence by a suitable subsequence Lemma 2.1 implies $\|f \otimes \psi - (f \otimes \widetilde{g_n})\mathbf{1}_V\|_{\mathcal{H}}^2 = \int_V (\psi - \widetilde{g_n})^2 \, d\Gamma(f) \to 0$ by bounded convergence, where $\widetilde{g_n} = \kappa_N \circ g_n$ with fixed $N \geq \|\psi\|_{\sup}$. □

Example 4

(1) In Example 2(1) the space $\mathcal{H}$ is isometrically isomorphic to the space $L_2(\mathbb{R}^n, \mathbb{R}^n)$ of $\mathbb{R}^n$-valued square integrable functions on $\mathbb{R}^n$.
(2) For the Riemannian situation in Example 2(2) the space $\mathcal{H}$ is isometrically isomorphic to the space $L_2(U, T^*M, dvol)$ of L_2-differential 1-forms on $U \subset M$.

Remark 2

(i) The spaces $\mathcal{H}_x$ may be seen as the *fibers* of the measurable L_2-bundle $\mathcal{H}$. Formula (4.3) expresses the *fibers* in terms of coordinates.
(ii) The spaces $\mathcal{H}_x$ depend on the choice of m. However, the space $\mathcal{H}$ does not, as follows from (2.2) and (4.5).
(ii) If the coordinates y^i have finite energy then we may replace $\mathcal{C}$ by $\mathcal{P}(y)$ in (4.5) and the subsequent formulas. By Lemma 4.2, regularity and [13, Theorem 2.1.4] this yields the same space $\mathcal{H}$.
(iii) We formulated (4.5) and (4.6) in terms of the algebra $\mathcal{C}$ in order to use the same definition of the space of 1-forms as in [7, 23, 25, 28]. Alternatively – and in view of Definition 3.1 this seems more appropriate – one can endow $\mathcal{P}(y) \otimes \tilde{\mathcal{P}}(y)$ with a directed family of Hilbert seminorms determined by $\|f \otimes g\|_{\mathcal{H}(V)} := \|(f \otimes g)\mathbf{1}_V\|_{\mathcal{H}}$, where the sets V are relatively compact and open. This yields a presheaf of Hilbert spaces whose inverse limit is a locally convex space $\mathcal{H}_{loc}$. Details can be found in [24, Section 6]. Also $\mathcal{F}_{loc}$ may be viewed as a locally convex space, and the derivation ∂ may then be interpreted as a continuous linear operator from $\mathcal{F}_{loc}$ into $\mathcal{H}_{loc}$, if (4.8) is replaced by $\|\partial f\|_{\mathcal{H}(V)}^2 = \Gamma(f)(V), f \in \mathcal{P}(V)$.

5 Differential and Gradient in Coordinates

For any coordinate function y^i and any relatively compact open set V the element $(\partial y^i)\mathbf{1}_V$ is an element of $\mathcal{H}$. This implies the identities

$$\left\langle (\partial y^i)_x, (\partial y^j)_x \right\rangle_{\mathcal{H}_x} = Z^{ij}(x)$$

for m-a.e. $x \in X$. Moreover, the local version of (4.7) shows that for any function $f = F(y^{n_1}, \dots, y^{n_k})$ from $\mathcal{P}(y)$ we have on any locally compact open set V

$$\partial f = \sum_{i=1}^{k} \frac{\partial F}{\partial y^{n_i}} \partial y^{n_i} \tag{5.1}$$

Example 5 In the Euclidean and Riemannian situations (1) and (2) in Examples 2 the operator ∂ may be identified with the exterior derivation and formula (5.1) becomes the classical identity in (4.4).

The operator ∂ may be viewed as a *generalization of the exterior derivation* and (5.1) may be viewed as a formula for the *differential* ∂f of f in terms of coordinates.

On a general metric measure space a smooth theory of ordinary differential equations is not available. On the other hand the spaces $\mathcal{H}_x$ are Hilbert, hence self-dual. Therefore it seems artificial to rigorously distinguish between 1-forms and vector fields. We interpret the elements of $\mathcal{H}$ also as *(measurable) vector fields* and ∂ as a *substitute for the gradient operator*.

Recall the notation in (3.1). Given a finite ordered subset J of I let the collection of $\mathcal{H}_x$-equivalence classes of elements of $\mathcal{P}_J \otimes \mathcal{P}_J(y)$ be denoted by $\mathcal{H}_{x,J}$. Clearly this is a subspace of $\mathcal{H}_x$, and we have

$$\mathcal{H}_x = \operatorname{clos}\left(\bigcup_{J \subset I} \mathcal{H}_{x,J} \right),$$

the union taken over all finite ordered subsets J of I.

Now suppose $J = (n_1, \dots, n_k)$. Formula (5.1) implies that the elements $(\partial y^{n_1})_x$, $\dots$, $(\partial y^{n_k})_x$ span $\mathcal{H}_{x,J}$. Let $Z_J(x)$ denote the matrix $(Z(x)^{n_i n_j})_{i,j=1}^k$, clearly symmetric and nonnegative definite. The preceding formulas yield another *expression of the gradient* ∂f*, now in terms of the Euclidean gradient and the measurable metric* Z: For any $f = F \circ y \in \mathcal{P}_J(y)$ and any $j = 1, \dots, k$ we have

$$\langle (\partial f)_x, (\partial y^{n_j})_x \rangle_{\mathcal{H}_x} = \sum_{i=1}^{k} \frac{\partial F}{\partial y^{n_i}}(y) Z^{n_i n_j}(x) = (Z_J(x) \nabla F(y))_j, \tag{5.2}$$

where ∇F is the gradient of F on $\mathbb{R}^k$.

Example 6

(1) For Examples 2 (1) we obtain

$$\left\langle (\partial f)_x, (\partial y^j)_x \right\rangle_{\mathcal{H}_x} = \sum_{i=1}^{n} a_{ij}(x) \frac{\partial f}{\partial y^i}(x) = (a(x) \nabla f(x))_j .$$

(2) In the Riemannian case of Examples 2 (2) formula (5.2) gives

$$\left\langle (\partial f)_x, (\partial y^j)_x \right\rangle_{\mathcal{H}_x} = \left\langle df, dy^j \right\rangle_{T_x^* M} = \frac{\partial f}{\partial y^i}(x) \left\langle dy^i, dy^j \right\rangle_{T_x^* M} = g^{ij}(x) \frac{\partial f}{\partial y^i}(x).$$

This equals $dy^j(\mathrm{grad} f)$ because $\mathrm{grad} f = g^{ij} \frac{\partial f}{\partial y^i} \frac{\partial}{\partial y^j}$.

(3) Let $\langle\langle \cdot, \cdot \rangle\rangle$ denote the cometric associated with the Heisenberg group $\mathbb{H}$. Then

$$\left\langle (\partial f)_q, (\partial y^1)_q \right\rangle_{\mathcal{H}_q} = \sum_{i=1}^{3} \frac{\partial f}{\partial y^i}(q) \left\langle\left\langle dy^i, dy^1 \right\rangle\right\rangle = (Z(q)\nabla f(q))_1 = Xf(q).$$

In a similar manner we obtain

$$\left\langle (\partial f)_q, (\partial y^2)_q \right\rangle_{\mathcal{H}_q} = (Z(q)\nabla f(q))_2 = Yf(q)$$

$$\left\langle (\partial f)_q, (\partial y^3)_q \right\rangle_{\mathcal{H}_q} = (Z(q)\nabla f(q))_3 = -\frac{\eta}{2} Xf(q) + \frac{\xi}{2} Yf(q).$$

6 Divergence in Coordinates

By $-\partial^*$ we denote the adjoint of ∂, that is the unbounded linear operator $-\partial^* : \mathcal{H} \to L_2(X, \mu)$ with dense domain $\mathrm{dom}\, \partial^*$ and such that the integration by parts formula

$$\langle v, \partial u \rangle_{\mathcal{H}} = -\left\langle \partial^* v, u \right\rangle_{L_2(X,\mu)} \tag{6.1}$$

holds for all $v \in \mathrm{dom}\, \partial^*$ and $f \in \mathcal{F}$. We view the operator $-\partial^*$ both ways, as *coderivation* and as *divergence operator*.

In the context of coordinates it is more suitable to deviate a bit from the Hilbert space interpretation in (6.1). First assume that all coordinates y^i have finite energy. For an element $(\partial f)g$ of $\mathcal{H}$ with $f, g \in \mathcal{P}(y)$ we then set

$$\partial^*((\partial f)g)(u) := -\left\langle (\partial f)g, (\partial u) \right\rangle_{\mathcal{H}}, \quad u \in \mathcal{P}(y).$$

By Cauchy-Schwarz $|\partial^*((\partial f)g)(u)| \le \|(\partial f)g\|_{\mathcal{H}}\ \mathcal{E}(u)$, and therefore $\partial^*(\partial f)g$ may be seen as a continuous linear functional on $\mathcal{P}(u)$, and after a straighforward extension by Definition 3.1 and regularity, on $\mathcal{F}$.

As before let $J = (n_1, \ldots, n_k)$. Given functions polynomials F and G in $y^{n_1}, \ldots, y^{n_k}$ and a function $u = U \circ y$ with $U \in C^1(\mathbb{R}^k)$ put

$$\mathrm{div}_{Z_J}(G\nabla F)(U) := -\sum_{i,j=1}^{k} \int_X G(y) \frac{\partial F}{\partial y^{n_i}}(y) Z^{n_i, n_j}(x) \frac{\partial U}{\partial y^{n_j}}(y) m(dx).$$

Then

$$\partial^*((\partial f)g)(u) = \mathrm{div}_{Z_J}(G\nabla F)(U) \tag{6.2}$$

provides a 'distributional' *coordinate expression for the divergence*. Of course this is a naive definition by duality, and in particular we have $\partial^*((\partial f))(u) = -\mathcal{E}(f,u)$. In general there is no integration by parts formula on the level of coordinates that could permit a more interesting definition.

If the coordinates y^i do not have finite energy, we view $\mathcal{P}(y)$ as a locally convex space, then $\partial^*(\partial f)g\mathbf{1}_V$ with relatively compact open V defines a continuous linear functional on $\mathcal{P}(y)$. Proceeding similarly as before one obtains local versions of (6.2).

Example 7

(1) For Example 2 (1) we obtain

$$\mathrm{div}_a(g\nabla f)(u) = -\sum_{i,j=1}^n \int_{\mathbb{R}^n} g(x)\frac{\partial f}{\partial x_i}(x)a_{ij}(x)\frac{\partial u}{\partial x_j}(x)dx$$

for any $u \in C_c^1(\mathbb{R}^n)$. If in addition the coefficients a_{ij} are C^1, this is seen to equal

$$\int_{\mathbb{R}^n} \mathrm{div}(a(g\nabla f))u\,dx.$$

(2) In the Riemannian situation of Examples 2 (ii) we have

$$\mathrm{div}_g(h\nabla f)(u) = \int_W g^{ij}h\frac{\partial f}{\partial y^i}\frac{\partial u}{\partial y^j}\sqrt{g}dy^1\cdots dy^n = \int_U \mathrm{div}(h\,\mathrm{grad} f)\,u\,dvol$$

for any $u \in C_c^1(U)$, where $g := \det(g_{ij})$ and

$$\mathrm{div}(h\,\mathrm{grad} f) = \frac{1}{\sqrt{g}}\frac{\partial}{\partial y^j}\left(\sqrt{g}g^{ij}h\frac{\partial f}{\partial y^i}\right)$$

is the divergence of $h\,\mathrm{grad} f$ in the usual Riemannian sense. See [29, Section 2.1].

(3) In Example 2 (3) formula (6.2) yields

$$\mathrm{div}_Z(g\nabla f)(u) = \sum_{i,j=1}^3 \int_U Z^{ij}g\frac{\partial f}{\partial y^i}\frac{\partial u}{\partial y^j}\,d\lambda^3 = \sum_{i,j=1}^3 \int_U \frac{\partial}{\partial y^j}\left(Z^{ij}g\frac{\partial f}{\partial y^i}\right)\,u\,d\lambda^3$$

for any $u \in C_c^1(U)$, what equals

$$\int_U \left(\frac{\partial}{\partial \xi} gXf + \frac{\partial}{\partial \eta} gYf + \frac{\partial}{\partial \zeta} \left(g(-\frac{\eta}{2}Xf + \frac{\xi}{2}Yf) \right) \right) u \, d\lambda^3 = \int_U \operatorname{div}(Z(g\nabla f)) \, u \, d\lambda^3,$$

where div is the ordinary divergence operator on $\mathbb{R}^3$.

7 Generator in Coordinates

We consider the infinitesimal generator $(L, \operatorname{dom} L)$ of $(\mathcal{E}, \mathcal{F})$. From (6.1) and the definition of the adjoint we see that for any $f \in \operatorname{dom} L$ we have $\partial f \in \operatorname{dom} \partial^*$ and

$$Lf = \partial^* \partial f. \tag{7.1}$$

Although in general a coordinate version of this formula may not be available, it can be written in terms of coordinates for specific examples.

To express L in coordinates additional assumptions are inevitable. Even if $y^i \in \operatorname{dom} L$ for all i the inclusion $\mathcal{P}(y) \subset \operatorname{dom} L$ holds if and only if the reference measure μ itself is energy dominant, that is if $(\mathcal{E}, \mathcal{F})$ admits a *carré du champ* in the sense of [5]. For Examples 2 (1)–(4) this is satisfied. However, the standard resistance form on the Sierpinski gasket, considered as a Dirichlet form with respect to the natural self-similar Hausdorff measure, does not have this property, and this situation is typical for a large class of self-similar spaces, [4, 14, 17].

Assumption 1 The reference measure μ itself is energy dominant.

Let $(L, \operatorname{dom}_{(1)} L)$ denote the smallest closed extension of the restriction of L to

$$\{f \in \operatorname{dom} L \cap L_1(X, \mu) : Lf \in L_1(X, \mu)\}.$$

Assumption 1 is known to be necessary and sufficient for $\operatorname{dom}_{(1)} L \cap L_\infty(X, \mu)$ to be an algebra under pointwise multiplication. If it is in force, then $f, g \in \operatorname{dom} L$ implies $fg \in \operatorname{dom}_{(1)} L$ and we have

$$\frac{d\Gamma(f, g)}{d\mu} = L(fg) - fLg - gLf, \tag{7.2}$$

see [5, Theorems I.4.2.1 and I.4.2.2]. To formulate local conditions on the coordinate functions we follow [37, Definition 4.2 (2)] and say that a function $f \in L_{2,loc}(X, \mu)$ *belongs to the strong local domain* $\operatorname{dom}_{loc} L$ *of* L if for any relatively compact open set V there exists some $u \in \mathcal{F}$ such that $f|_V = u|_V$ μ-a.e. Similarly we define $\operatorname{dom}_{(1),\,loc} L$. Then identity (7.2) holds for any $f, g \in \operatorname{dom}_{loc} L$ locally on any relatively compact open set V.

Assumption 2 The coordinates y^i are members of $\mathrm{dom}_{loc}\ L$.

Let Assumptions 1 and 2 be in force. This implies $\mathcal{P}(y) \subset \mathrm{dom}_{loc}\ L$. Suppose $f = F \circ y \in \mathcal{P}(y)$, where again $J = (n_1, \ldots, n_k)$. Using (7.2) on the coordinates y^i and iterating, we inductively arrive at a *coordinate formula for the generator*

$$Lf(x) = \sum_{i,j=1}^{k} \frac{\partial^2 F}{\partial y^{n_j} \partial y^{n_j}}(y) Z^{n_i n_j}(x) + \sum_{i=1}^{k} \frac{\partial F}{\partial y^{n_i}}(y) L y^{n_i}(x),$$

valid locally on any relatively compact open V. This is a version of a well known identity, see e.g. [12, Lemma 6.1] or [11].

Example 8

(1) For Example 2 (1) with C^1-coefficients a_{ij} we have

$$Lf = \mathrm{div}(a\nabla f) = \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j} a_{ij} + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} \sum_{j=1}^{n} \frac{\partial a_{ij}}{\partial x_j}.$$

(2) For Example 2 (2) we observe

$$\Delta f = \mathrm{div}(\mathrm{grad} f) = \frac{1}{\sqrt{g}} \frac{\partial}{\partial y^j} \left(\sqrt{g} g^{ij} \frac{\partial f}{\partial y^i} \right),$$

what differs by a minus sign from the Laplace-Beltrami operator (convention).

(3) For Example 2 (3) arrive at the Heisenberg sub-Laplacian,

$$Lf = \mathrm{div}(Z\nabla f) = \frac{\partial^2 f}{\partial \xi^2} + \frac{\partial^2 f}{\partial \eta^2} + \xi \frac{\partial^2 f}{\partial \eta \partial \zeta} + \eta \frac{\partial^2 f}{\partial \xi \partial \zeta} + \frac{\xi^2 + \eta^2}{4} \frac{\partial^2 f}{\partial \zeta^2} = \left(X^2 + Y^2\right) f.$$

(4) In Example 2 (4) the Dirichlet form generator of $(\mathcal{E}, \mathcal{F})$ on $L_2(K, \nu)$ is the Kusuoka Laplacian $(\Delta_\nu, \mathrm{dom}\ \Delta_\nu)$. The coordinate functions y^i are harmonic, that is $y^i \in \mathrm{dom}\ \Delta_\nu$ and $\Delta_\nu y^i = 0$, $i = 1, 2$. Accordingly we have

$$\Delta_\nu f(x) = \sum_{i,j=1}^{2} \frac{\partial^2 F}{\partial y^i \partial y^j}(y) Z^{ij}(x)$$

for any $f = F \circ y \in \mathcal{P}(y)$. This can be rewritten as $\mathrm{tr}(Z(x) D^2 F(y))$, where $D^2 F$ is the Hessian of F and tr the trace operator, see [45, Theorem 8].

8 Constructing Coordinate Sequences

Let $(\mathcal{E},\mathcal{F})$ be a strongly local regular Dirichlet form. Under some continuity condition it is always possible to simultaneously construct an energy dominant measure and a corresponding coordinate sequence. The latter may be designed to have nice decay properties. Let $(P_t)_{t>0}$ denote the Markovian semigroup uniquely associated with $(\mathcal{E},\mathcal{F})$, [6, 13]. If it is also a strongly continuous semigroup of contractions $P_t : C_0(X) \to C_0(X)$ on the space $C_0(X)$ of continuous functions vanishing at infinity, then it is called a *Feller semigroup*.

Example 9 The transition semigroups of many diffusion processes of Euclidean domains or manifolds are Feller semigroups. Also the semigroups of many diffusions on fractals are known to be Feller, see for instance [1–3, 33].

Lemma 8.1 *Assume that the semigroup $(P_t)_{t>0}$ is a Feller semigroup. Then there exist a finite energy dominant measure $\tilde{m}$ and a coordinate sequence $(y^i)_{i\in I} \subset \operatorname{dom} L$ for $(\mathcal{E},\mathcal{F})$ with respect to $\tilde{m}$ such that*

(i) $\operatorname{span}(\{y^i\}_{i\in I})$ *is dense in $\mathcal{F}$,*
(ii) *For any i also the functions Ly^i are continuous,*
(iii) *We have* $\sum_{i=1}^{\infty} \|y^i\|_{\sup}^2 < +\infty$ *and* $\sum_{i=1}^{\infty} \|Ly^i\|_{\sup}^2 < +\infty.$

Proof Let $\{f_i\}_i \subset C_c(X)$ be a countable family of nonzero functions that is dense in $L_2(X,\mu)$. By the Feller property, the resolvent functions $G_1 f_i(x) := \int_0^\infty e^{-t} P_t f(x)dt$, are continuous and $G_1 f_i \in \operatorname{dom} L$. Set

$$y^i := 2^{-n} G_1 f_i \Big/ \left(\|G_1 f_i\|_{\sup} + \|f_i\|_{\sup} + \mathcal{E}(G_1 f_i)^{1/2}\right).$$

Then (ii) and (iii) are satisfied. The range $\operatorname{Im} G_1$ of $G_1 : L_2(X,\mu) \to L_2(X,\mu)$ is dense in $\mathcal{F}$ and any element of $\operatorname{Im} G_1$ can be approximated in $\mathcal{F}$ by linear combinations of the functions $G_1 f_i$, what implies (i). Now set $\tilde{m} := \sum_{i=1}^{\infty} 2^i \Gamma(y^i)$. Because the energy measures $\Gamma(y^i)$ satisfy $\Gamma(y^i) \le \dfrac{\Gamma(G_1 f_i)}{2^{2n}\mathcal{E}(G_1 f_i)} \le 2^{-2i}$, we have $\tilde{m}(X) \le \sum_{i=1}^{\infty} 2^{-i} < +\infty$. For the densities we observe $Z^{ii} = \frac{d\Gamma(y^i)}{d\tilde{m}} \le \frac{d\Gamma(y^i)}{2^i d\Gamma(y^i)} \le 2^{-i}$ $\tilde{\mathrm{m}}-$a.e. Polarizing and choosing appropriate $\tilde{m}$-versions of the functions Z^{ij}, we may assume that for m-a.e. $x \in X$ and any $N \in \mathbb{N}$ the matrix $(Z^{ij}(x))_{i,j=1}^N$ is symmetric and nonnegative definite. To do so it suffices to note that given $v_1,\ldots,v_N \in \mathbb{R}$, $0 \le \Gamma\left(\sum_{i=1}^N v_i y^i\right)(A) = \int_A \sum_{i=1}^N Z^{ij}(x) v_i v_j \, \tilde{m}(dx)$ is a nonnegative Radon measure, hence its density must be nonnegative $\tilde{m}$-a.e. By letting N go to infinity we can finally obtain

$$\|Z(x)v\|_{l_2}^2 \le \sum_{i,j} |Z^{ij}(x)|^2 |v_j|^2 \le \sum_{i,j} |Z^{ii}(x)||Z^{jj}(x)||v_j|^2 \le \sum_{i,j} 2^{-i-j}|v_j|^2 \le \|v\|_{l_2}^2$$

for any $v = (v_1, v_2, \dots) \in l_2$, what allows to conclude that $Z(x)$ is bounded, symmetric and nonnegative definite on l_2 for μ-a.e. $x \in X$. □

References

1. M.T. Barlow, *Diffusions on fractals*. Lectures on Probability Theory and Statistics (Saint-Flour, 1995), 1–121, Lecture Notes in Math., 1690 (Springer, Berlin, 1998)
2. M.T. Barlow, R.F. Bass, Brownian motion and harmonic analysis on Sierpinski carpets. Canad. J. Math. **51**, 673–744 (1999)
3. M.T. Barlow, R.F. Bass, T. Kumagai, A. Teplyaev, Uniqueness of Brownian motion on Sierpinski carpets. J. Eur. Math. Soc. **12**, 655–701 (2010)
4. O. Ben-Bassat, R.S. Strichartz, A. Teplyaev, What is not in the domain of the Laplacian on a Sierpinski gasket type fractal. J. Funct. Anal. **166**, 197–217 (1999)
5. N. Bouleau, F. Hirsch, *Dirichlet Forms and Analysis on Wiener Space*. deGruyter Studies in Mathematics, vol. 14 (deGruyter, Berlin, 1991)
6. Z.-Q. Chen, M. Fukushima, *Symmetric Markov Processes, Time Change, and Boundary Theory* (Princeton University Press, Princeton, 2012)
7. F. Cipriani, J.-L. Sauvageot, Derivations as square roots of Dirichlet forms. J. Funct. Anal. **201**, 78–120 (2003)
8. F. Cipriani, J.-L. Sauvageot, Fredholm modules on p.c.f. self-similar fractals and their conformal geometry. Commun. Math. Phys. **286**, 541–558 (2009)
9. F. Cipriani, D. Guido, T. Isola, J.-L. Sauvageot, Spectral triples for the Sierpinski gasket. J. Funct. Anal. **266**, 4809–4869 (2014)
10. B.K. Driver, T. Melcher, Hypoelliptic heat kernel inequalities on the Heisenberg group. J. Funct. Anal. **221**, 340–365 (2005)
11. A. Eberle, *Uniqueness and Non-uniqueness of Semigroups Generated by Singular Diffusion Operators*. Springer LNM, vol. 1718 (Springer, New York, 1999)
12. M. Emery, *Stochastic Calculus in Manifolds*. Springer Universitext (Springer, Berlin/Heidelberg, 1989)
13. M. Fukushima, Y. Oshima, M. Takeda, *Dirichlet Forms and Symmetric Markov Processes* (deGruyter, Berlin/New York, 1994)
14. M. Hino, On singularity of energy measures on self-similar sets. Probab. Theory Relat. Fields **132**, 265–290 (2005)
15. M. Hino, Measurable Riemannian structures associated with strong local Dirichlet forms. Math. Nachr. **286**, 1466–1478 (2013)
16. M. Hino, Energy measures and indices of Dirichlet forms, with applications to derivatives on some fractals. Proc. Lond. Math. Soc. **100**(1), 269–302 (2013)
17. M. Hino, K. Nakahara, On singularity of energy measures on self-similar sets II. Bull. Lond. Math. Soc. **38**, 1019–1032 (2006)
18. M. Hinz, *Sup-norm-closable bilinear forms and Lagrangians* (Preprint). arXiv:1407.1301
19. M. Hinz, *Magnetic energies and Feynman-Kac-Ito formulas for symmetric Markov processes* (Preprint). arXiv:1409.7743
20. M. Hinz, L. Rogers, *Magnetic fields on resistance spaces* (Preprint). arXiv:1501.01100
21. M. Hinz, A. Teplyaev, Dirac and magnetic Schrödinger operators on fractals. J. Funct. Anal. **265**(11), 2830–2854 (2013)
22. M. Hinz, A. Teplyaev, Local Dirichlet forms, Hodge theory, and the Navier-Stokes equations on topologically one-dimensional fractals. Trans. Am. Math. Soc. **367**, 1347–1380 (2015)
23. M. Hinz, D. Kelleher, A. Teplyaev, Measures and Dirichlet forms under the Gelfand transform, *Probability and statistics* **18**, Zapiski Nauchnyh Seminarov POMI **408**, 303–322 (2012). arXiv:1212.1099

24. M. Hinz, D. Kelleher, A. Teplyaev, Metrics and spectral triples for Dirichlet and forms. J. Noncommut. Geom. (To appear). arXiv:1309.5937
25. M. Hinz, M. Röckner, A. Teplyaev, Vector analysis for Dirichlet forms and quasilinear PDE and SPDE on metric measure spaces. Stoch. Proc. Appl. **123**(12), 4373–4406 (2013)
26. S. Gallot, D. Hulin, J. Lafontaine, *Riemannian Geometry* (Springer, Berlin/Heidelberg, 1990)
27. M. Gordina, Th. Laetsch, *Weak convergence to Brownian motion on sub-Riemannian manifolds* (2014, preprint). arXiv:1403.0142
28. M. Ionescu, L. Rogers, A. Teplyaev, Derivations and Dirichlet forms on fractals. J. Funct. Anal. **263**(8), 2141–2169 (2012)
29. J. Jost, *Riemannian Geometry and Geometric Analysis* (Springer, Heidelberg, 2002)
30. N. Kajino, Heat kernel asymptotics for the measurable Riemannian structure on the Sierpinski gasket. Potential Anal. **36**, 67–115 (2012)
31. N. Kajino, Time changes of local Dirichlet spaces by energy measures of harmonic functions. Forum Math. **24**, 339–363 (2012)
32. J. Kigami, *Harmonic metric and Dirichlet form on the Sierpiński gasket.* Asymptotic problems in probability theory: stochastic models and diffusions on fractals (Sanda/Kyoto, 1990), 201–218, Pitman Res. Notes Math. Ser., **283**, (Longman Sci. Tech., Harlow, 1993)
33. J. Kigami, *Analysis on Fractals* (Cambridge University Press, Cambridge, 2001)
34. J. Kigami, Measurable Riemannian geometry on the Sierpinski gasket: the Kusuoka measure and the Gaussian heat kernel estimate. Math. Ann. **340**(4), 781–804 (2008)
35. S. Kusuoka, Dirichlet forms on fractals and products of random matrices. Publ. Res. Inst. Math. Sci. **25**, 659–680 (1989)
36. S. Kusuoka, Lecture on diffusion process on nested fractals, in *Statistical Mechanics and Fractals*. Lecture Notes in Mathematics, vol. 1567 (Springer, Berlin, 1993), pp. 39–98
37. D. Lenz, A. Teplyaev, Expansion in generalized eigenfunctions for Laplacians on graphs and metric measure spaces. Trans. Am. Math. Soc. (To appear). arXiv:1310.5650
38. R. Montgomery, *A Tour of Sub-Riemannian Geometries, Their Geodesics and Applications*. Mathematical Surveys and Monographs, vol. 91 (American Mathematical Society, Providence, 2002)
39. S. Nakao, Stochastic calculus for continuous additive functionals. Z. Wahrsch. verw. Geb. **68**, 557–578 (1985)
40. J.-L. Sauvageot, Quantum differential forms, differential calculus and semigroups, in *Quantum Probability and Applications V*. Lecture Notes in Mathematics, vol. 1442 (Springer, New York, 1990), pp. 334–346
41. R.S. Strichartz, Sub-Riemannian geometry. J. Diff. Geom. **24**(2), 221–263 (1986)
42. R.S. Strichartz, Taylor approximations on Sierpinski type fractals. J. Funct. Anal. **174**, 76–127 (2000)
43. R.S. Strichartz, *Differential Equations on Fractals: A Tutorial* (Princeton University Press, Princeton, 2006)
44. A. Teplyaev, Gradients on fractals. J. Funct. Anal. **174**, 128–154 (2000)
45. A. Teplyaev, Harmonic coordinates on fractals with finitely ramified cell structure. Canad. J. Math. **60**, 457–480 (2008)

Fractal Zeta Functions and Complex Dimensions: A General Higher-Dimensional Theory

Michel L. Lapidus, Goran Radunović, and Darko Žubrinić

Abstract In 2009, the first author introduced a class of zeta functions, called 'distance zeta functions', which has enabled us to extend the existing theory of zeta functions of fractal strings and sprays (initiated by the first author and his collaborators in the early 1990s) to arbitrary bounded (fractal) sets in Euclidean spaces of any dimensions. A closely related tool is the class of 'tube zeta functions', defined using the tube function of a fractal set. These zeta functions exhibit deep connections with Minkowski contents and upper box (or Minkowski) dimensions, as well as, more generally, with the complex dimensions of fractal sets. In particular, the abscissa of (Lebesgue, i.e., absolute) convergence of the distance zeta function coincides with the upper box dimension of a set. We also introduce a class of transcendentally quasiperiodic sets, and describe their construction based on a sequence of carefully chosen generalized Cantor sets with two auxilliary parameters. As a result, we obtain a family of "maximally hyperfractal" compact sets and relative fractal drums (i.e., such that the associated fractal zeta functions have a singularity at every point of the critical line of convergence). Finally, we discuss the general fractal tube formulas and the Minkowski measurability criterion obtained by the authors in the context of relative fractal drums (and, in particular, of bounded subsets of $\mathbb{R}^N$).

The research of Michel L. Lapidus was partially supported by the National Science Foundation under grants DMS-0707524 and DMS-1107750, as well as by the Institut des Hautes Études Scientifiques (IHÉS) where the first author was a visiting professor in the Spring of 2012 while part of this research was completed. This paper is based on a plenary lecture given by the first author (MLL) at the International Conference on "Fractal Geometry and Stochastics" held in Tabarz, Germany, in March 2014. Goran Radunović and Darko Žubrinić express their gratitude to the Ministry of Science of the Republic of Croatia for its support.

M.L. Lapidus (✉)
Department of Mathematics, University of California, Riverside, CA 92521-0135, USA
e-mail: lapidus@math.ucr.edu

Goran Radunović • Darko Žubrinić
Faculty of Electrical Engineering and Computing, Department of Applied Mathematics, University of Zagreb, Unska 3, 10000 Zagreb, Croatia
e-mail: goran.radunovic@fer.hr; darko.zubrinic@fer.hr

C. Bandt et al. (eds.), *Fractal Geometry and Stochastics V*, Progress in Probability 70, DOI 10.1007/978-3-319-18660-3_13

Keywords Fractal set • Fractal zeta functions • Distance zeta function • Tube zeta function • Geometric zeta function of a fractal string • Minkowski content • Minkowski measurability • Upper box (or Minkowski) dimension • Complex dimensions of a fractal set • Holomorphic and meromorphic functions • Abscissa of convergence • Quasiperiodic function • Quasiperiodic set • Relative fractal drum • Fractal tube formulas

1 Introduction

This article provides a short survey of some of the recent advances in the theory of fractal zeta functions and the associated higher-dimensional theory of complex dimensions, valid for arbitrary bounded subsets of Euclidean spaces and developed in the forthcoming research monograph [41], entitled *Fractal Zeta Functions and Fractal Drums*: *Higher-Dimensional Theory of Fractal Dimensions*. (See also the research articles [42–46] and the survey article [47].)

The theory of zeta functions of fractal strings, initiated by the first author in the early 1990s and described in an extensive research monograph [52], joint with M. van Frankenhuijsen (see also the references therein), was given an unexpected impetus in 2009, when a new class of zeta functions, called 'distance zeta functions', was discovered (also by the first author).[1] Since distance zeta functions are associated with arbitrary bounded (fractal) sets in Euclidean spaces of any dimension (see Definition 2.1), they clearly represent a valuable tool connecting the geometry of fractal sets with complex analysis. This interplay is described in [41–47], where the foundations of the theory of fractal zeta functions have been laid. In this paper, by 'fractal zeta functions' we mean the following three classes of zeta functions: zeta functions of fractal strings (and, more generally, of fractal sprays), distance zeta functions and tube zeta functions of bounded subsets of $\mathbb{R}^N$, with $N \geq 1$, although some other classes may appear as well, like zeta functions of relative fractal drums in $\mathbb{R}^N$ and spectral zeta functions; see Sect. 6 below and [41, Chap. 4]. The theory of fractal zeta functions exhibits very interesting connections with the Minkowski contents and dimensions of fractal sets; see Theorems 2.3 and 2.5.

Like fractal string theory, which the present theory of fractal zeta functions extends to arbitrary dimensions (as well as to "relative fractal drums" in $\mathbb{R}^N$), the work described here should eventually have applications to various aspects of harmonic analysis, fractal geometry, dynamical systems, geometric measure theory and analysis on nonsmooth spaces, number theory and arithmetic geometry, mathematical physics and, more speculatively, to aspects of condensed matter

[1] For fractal string theory and the associated one-dimensional theory of complex dimensions, as well as for the extensions to higher-dimensional fractal sprays (in the sense of [39]), we refer the reader to the research monographs [50–52] along, for example, with the articles [5, 6, 12, 14–16, 20–40, 48, 49, 53, 56, 60, 61]. We refer, in particular, to [52, §12.2.1 and Chap. 13] for a survey of some of the recent developments of the theory, prior to [41–47].

physics and cosmology. Some of the more mathematical applications of the theory are described in [41], as well as in [42–47], but a variety of potential applications remain to be explored or even imagined.

The basic property of the distance zeta function of a fractal set, described in Theorem 2.2, is that its abscissa of (absolute or Lebesgue) convergence is equal to the upper box dimension D of the set. Under some mild hypotheses, D is always a singularity; see part (*b*) of Theorem 2.2. Furthermore, assuming that D is a pole, then it is simple. Moreover, the residue of the distance zeta function computed at D is, up to a multiplicative constant, between the corresponding upper and lower Minkowski contents. A similar statement holds for the tube zeta function. (See Theorems 2.3 and 2.5, respectively.)

In addition, according to part (*b*) of Theorem 2.2, under some mild assumptions on a bounded set A, the abscissa of (Lebesgue, i.e., *absolute*) convergence of its distance zeta function coincides not only with D, but also with the abscissa of *holomorphic continuation* of the zeta function.

We stress that if $D := \overline{\dim}_B A < N$, all the results concerning the distance zeta functions have exact counterparts for the tube zeta functions, and vice versa. In other words, the fractal zeta functions introduced in [41–47] contain essentially the same information. In practice, however, it is often the case that one of the fractal zeta functions is better suited for the given situation under consideration.

In Sect. 3, we discuss the existence and the construction of a suitable meromorphic continuation of the distance (or tube) zeta function of a fractal set, both in the Minkowski measurable case (Theorem 3.1) and a frequently encountered instance of Minkowski nonmeasurable case (Theorem 3.2). We will illustrate the latter situation by computing the fractal zeta function and the associated complex dimensions of the Sierpiński carpet; see Proposition 3.3 and Example 4 when $N = 2$ or 3, respectively. Many other examples are provided in [41] and [42–46], where are calculated, in particular, the complex dimensions of the higher-dimensional Sierpiński gaskets and carpets in $\mathbb{R}^N$, for any $N \geq 2$.

In Sect. 4, we introduce the so-called transcendentally n-quasiperiodic sets, for any integer $n \geq 2$ (that is, roughly speaking, the sets possessing n quasiperiods; see Definition 4.4), and describe the construction of 2-quasiperiodic sets, based on carefully chosen generalized Cantor sets with two parameters, introduced in Definition 4.1; see Theorem 4.5. It is also possible to construct n-quasiperiodic sets, for any $n \geq 2$, and even ∞-quasiperiodic sets, that is, sets which possess infinitely many quasiperiods; see Sect. 5 below and [41, §4.6].

In Sect. 6, we introduce the notion of a relative fractal drum (A, Ω) (which represents a natural extension of the notion of bounded fractal string and of bounded set). We also introduce the corresponding relative distance and tube zeta functions $\zeta_A(\,\cdot\,, \Omega)$ and $\tilde{\zeta}_A(\,\cdot\,, \Omega)$, and study their properties. It is noteworthy that the relative box dimension $\dim_B(A, \Omega)$ can be naturally defined as a real number, which may also assume negative values, including $-\infty$.

In Sect. 7, we address the question of reconstructing the tube function $t \mapsto |A_t \cap \Omega|$ of a relative fractal drum (A, Ω), and thereby of obtaining a general "fractal tube formula" expressed in terms of the complex dimensions of (A, Ω) (defined as the poles of a suitable meromorphic extension of the relative distance zeta function $\zeta_A(\cdot\,,\Omega)$). The corresponding tube formulas are obtained in [41, Chap. 5] and [46] (announced in [45]), as well as illustrated by a variety of examples. The example of the three-dimensional Sierpiński carpet is given in Example 4. Moreover, towards the end of Sect. 7, we explain how to deduce from our general tube formulas (and significantly extend) earlier results obtained for fractal strings (in [50–52]) and, especially, for fractal sprays and self-similar tilings (in [35] and [36]).

In closing this introduction, we recall some basic notation and terminology which will be needed in the sequel. First of all, in order to avoid trivial special cases, we assume implicitly that all bounded subsets of $\mathbb{R}^N$ under consideration in the statements of the theorems are nonempty. Assume that A is a given bounded subset of $\mathbb{R}^N$ and let r be a fixed real number. We define the *upper* and *lower r-dimensional Minkowski contents* of A, respectively, by

$$\mathcal{M}^{*r}(A) := \limsup_{t\to 0^+}\frac{|A_t|}{t^{N-r}},\quad \mathcal{M}_*^r(A) := \liminf_{t\to 0^+}\frac{|A_t|}{t^{N-r}},$$

where A_t denotes the Euclidean t-neighborhood of A (namely, $A_t := \{x \in \mathbb{R}^N : d(x,A) < t\}$) and $|A_t|$ is the N-dimensional Lebesgue measure of A_t. The *upper* and *lower box* (or *Minkowski*) *dimensions* of A are then defined, respectively, by

$$\overline{\dim}_B A := \inf\{r\in\mathbb{R} : \mathcal{M}^{*r}(A) = 0\},\quad \underline{\dim}_B A := \inf\{r\in\mathbb{R} : \mathcal{M}_*^r(A) = 0\}.$$

It is easy to check that $0 \le \underline{\dim}_B A \le \overline{\dim}_B A \le N$. Furthermore, if A is such that $\underline{\dim}_B A = \overline{\dim}_B A$, then this common value is denoted by $\dim_B A$ and is called the *box* (or *Minkowski*) *dimension* of A. Moreover, if A is such that, for some $D \in [0,N]$, we have $0 < \mathcal{M}_*^D(A) \le \mathcal{M}^{*D}(A) < \infty$ (in particular, then $\dim_B A$ exists and $D = \dim_B A$), we say that *A is Minkowski nondegenerate*. If $\mathcal{M}_*^D(A) = \mathcal{M}^{*D}(A)$, then this common value is denoted by $\mathcal{M}^D(A)$ and called the *Minkowski content* of A. Finally, assuming that A is such that $\mathcal{M}^D(A)$ exists and $0 < \mathcal{M}^D(A) < \infty$, we say that *$A$ is Minkowski measurable*.[2]

Throughout this paper, given $\alpha \in \mathbb{R}\cup\{\pm\infty\}$, we denote by $\{\operatorname{Re} s > \alpha\}$ the corresponding open right half-plane in the complex plane, defined by $\{s \in \mathbb{C} : \operatorname{Re} s > \alpha\}$. (In particular, if $\alpha = \pm\infty$, $\{\operatorname{Re} s > \alpha\}$ is equal to $\emptyset$ or $\mathbb{C}$, respectively.) Similarly, given any $\alpha \in \mathbb{R}$, we denote by $\{\operatorname{Re} s = \alpha\}$ the corresponding vertical line $\{s \in \mathbb{C} : \operatorname{Re} s = \alpha\}$.

[2] We note that the notion of Minkowski dimension was introduced (for noninteger values) by Bouligand [3] in the late 1920s (without making a clear distinction between the lower and upper limits), while the notions of (lower and upper) Minkowski content, Minkowski measurability and Minkowski nondegeneracy were introduced, respectively, in [9, 59] and [67]. (See also [22, 24, 38] and, especially, [39, 40], along with [52], for the latter notions.) For general references on the notion of Minkowski (or box) dimension (from different points of view), we refer, for example, to [7, 9, 54, 63] and [52].

2 Distance and Tube Zeta Functions

Let us introduce a new class of zeta functions, defined by the first author in 2009, which extends the notion of geometric zeta functions of bounded fractal strings to bounded subsets of Euclidean spaces of arbitrary dimensions.

Definition 2.1 ([41, 42]) Let A be a bounded subset of $\mathbb{R}^N$ and let δ be a fixed positive real number. Then, the *distance zeta function* ζ_A of A is defined by

$$\zeta_A(s) := \int_{A_\delta} d(x,A)^{s-N} \mathrm{d}x, \tag{2.1}$$

for all $s \in \mathbb{C}$ with $\mathrm{Re}\, s$ sufficiently large. Here, $d(x,A) := \inf\{|x-y| : y \in A\}$ denotes the usual Euclidean distance from x to A. Furthermore, the integral is taken in the sense of Lebesgue, and hence, is absolutely convergent.[3]

Remark 1 Since the difference of any two distance zeta functions of the same set A corresponding to two different values of δ is an entire function,[4] it follows that the dependence of the distance zeta function ζ_A on $\delta > 0$ is inessential, in the sense that the poles (of meromorphic extensions) of ζ_A, as well as their multiplicities, do not depend on the choice of δ.

The key for understanding the behavior of the distance zeta function ζ_A consists in understanding the Lebesgue integrability of the function $A_\delta \ni x \mapsto d(x,A)^{\mathrm{Re}\, s-N}$, where $s \in \mathbb{C}$ is fixed.[5] (We shall soon see that $\mathrm{Re}\, s$ should be sufficiently large.) More precisely, we are interested in the Lebesgue integrability of the function $x \mapsto d(x,A)^{-\gamma}$ defined on A_δ, where $\gamma := N - \mathrm{Re}\, s$ and s is a fixed complex number. Since the function is clearly bounded (and hence, integrable) for $\gamma \le 0$, it suffices to consider the case when $\gamma > 0$, that is, when $\mathrm{Re}\, s < N$.

Let us recall a useful and little known result due to Harvey and Polking, stated implicitly on page 42 of [13], in which a sufficient condition for Lebesgue integrability is expressed in terms of the upper box dimension. *If A is any nonempty bounded subset of $\mathbb{R}^N$, then the following implication holds*[6]:

$$\gamma < N - \overline{\dim}_B A \quad \Longrightarrow \quad \int_{A_\delta} d(x,A)^{-\gamma} \mathrm{d}x < \infty. \tag{2.2}$$

[3]For simplicity, we implicitly assume throughout this paper that $|A| = 0$; the case when $|A| > 0$ is discussed in [41].

[4]This is an easy consequence of the fact that $d(x,A) \in [\delta_1, \delta_2]$ for all $x \in A_{\delta_2} \setminus A_{\delta_1}$ with $0 < \delta_1 < \delta_2 < \infty$.

[5]Indeed, note that $|d(x,A)^{s-N}| = d(x,A)^{\mathrm{Re}\, s-N}$ for all $x \in A_\delta$.

[6]Moreover, if we assume that $D := \dim_B A$ exists, $D < N$ and $\mathcal{M}_*^D(A) > 0$, then the converse implication holds as well; see [67, Thm. 4.3]. (See also [68, Thm. 4.1(b)].)

Remark 2 The sufficient condition for the Lebesgue (i.e., absolute) integrability of the function $A_\delta \ni x \mapsto d(x, A)^{s-N}$ in the Harvey–Polking result in (2.2), becomes $\gamma := N - \operatorname{Re} s < N - \overline{\dim}_B A$, that is, $\operatorname{Re} s > \overline{\dim}_B A$. In other words, $\zeta_A(s)$ is well defined for all $s \in \mathbb{C}$ in the open right half-plane $\{\operatorname{Re} s > \overline{\dim}_B A\}$.

The distance zeta function of a bounded set represents a natural extension of the notion of *geometric zeta function* $\zeta_{\mathcal{L}}$, associated with a bounded fractal string $\mathcal{L} = (\ell_j)_{j\geq 1}$ (introduced by the first author and his collaborators[7] in the early 1990s and extensively studied in [50–52] and the relevant references therein):

$$\zeta_{\mathcal{L}}(s) := \sum_{j=1}^{\infty} (\ell_j)^s, \tag{2.3}$$

for all $s \in \mathbb{C}$ with $\operatorname{Re} s$ sufficiently large. Here, a *bounded fractal string* $\mathcal{L}$ is defined as a nonincreasing infinite sequence of positive real numbers $(\ell_j)_{j\geq 1}$ such that $\ell := \sum_{j\geq 1} \ell_j < \infty$. Alternatively, $\mathcal{L}$ can be viewed as a bounded open subset Ω of $\mathbb{R}$, in which case the ℓ_js are the lengths of the connected components (open intervals) of Ω, written in nonincreasing order (so that $\ell_j \downarrow 0$ as $j \to \infty$).

An important first result concerning $\zeta_{\mathcal{L}}$ (first observed in [23, 24], using a result from [2]) is that its abscissa of (absolute) convergence coincides with $\tilde{D}$ (the inner Minkowski dimension of $\mathcal{L}$ or, equivalently, of its fractal boundary $\partial\Omega$), defined by $\tilde{D} := \overline{\dim}_B(\partial\Omega, \Omega)$; see definition (6.1) below. For a direct proof of this statement, see [52, Thm. 1.10] or [52, Thm. 13.111] and [31]. In light of the next comment, it can be readily shown that part (a) of Theorem 2.2 below extends this result to arbitrary compact subsets of Euclidean spaces in any dimension; see [41, 42].

It is easy to see that the distance zeta function $\zeta_{A_{\mathcal{L}}}$ of the set

$$A_{\mathcal{L}} := \left\{ a_k := \sum_{j=k}^{\infty} \ell_j : k \geq 1 \right\} \subseteq [0, \ell],$$

associated with $\mathcal{L}$, and the geometric zeta function $\zeta_{\mathcal{L}}$ are connected by the following simple relation:

$$\zeta_{A_{\mathcal{L}}}(s) = u(s)\, \zeta_{\mathcal{L}}(s) + v(s), \tag{2.4}$$

for all complex numbers s such that $\operatorname{Re} s$ is sufficiently large, where u and v are holomorphic on $\mathbb{C}\setminus\{0\}$ and u is nowhere vanishing. In particular, due to Theorem 2.2 below, it follows that the abscissae of convergence of the distance zeta function ζ_A and of the geometric zeta function $\zeta_{\mathcal{L}}$ coincide, and that the corresponding poles located on the critical line $\{\operatorname{Re} s = \overline{\dim}_B A_{\mathcal{L}}\}$ (called *principal complex dimensions* of $\mathcal{L}$ or, equivalently, of $A_{\mathcal{L}}$), as well as their multiplicities, also coincide. The exact

[7] See, especially, [23, 24, 32, 33, 38–40] and [14].

same results hold if $A_{\mathcal{L}}$ is replaced by $\partial\Omega$, the boundary of Ω, where Ω is any geometric realization of $\mathcal{L}$ by a bounded open subset of $\mathbb{R}$. For more details, see [41, 42].

Before stating Theorem 2.2, we need to introduce some terminology and notation, which will also be used in the remainder of the paper.

Given a meromorphic function (or, more generally, an arbitrary complex-valued function) $f = f(s)$, initially defined on some domain $U \subseteq \mathbb{C}$, we denote by $D_{\mathrm{hol}}(f)$ the unique extended real number (i.e., $D_{\mathrm{hol}}(f) \in \mathbb{R} \cup \{\pm\infty\}$) such that $\{\mathrm{Re}\, s > D_{\mathrm{hol}}(f)\}$ is the *maximal* open right half-plane (of the form $\{\mathrm{Re}\, s > \alpha\}$, for some $\alpha \in \mathbb{R} \cup \{\pm\infty\}$) to which the function f can be holomorphically extended.[8] This maximal (i.e., largest) half-plane is denoted by $\mathcal{H}(f)$ and called the *half-plane of holomorphic continuation* of f.

If, in addition, the function $f = f(s)$ is assumed to be given by a *tamed Dirichlet-type integral* (or DTI, in short),[9] of the form

$$f(s) := \int_E \varphi(x)^s \mathrm{d}\mu(x), \tag{2.5}$$

for all $s \in \mathbb{C}$ with $\mathrm{Re}\, s$ sufficiently large, where μ is a (positive or complex) local (i.e., locally bounded) Borel measure on a given (measurable) space E and

$$0 \leq \varphi(x) \leq C \quad \text{for} \quad |\mu|\text{-a.e. } x \in E, \tag{2.6}$$

where $C \geq 0$,[10] then $D(f)$, the *abscissa of* (*absolute* or *Lebesgue*) *convergence* of f, is defined as the unique extended real number (i.e., $D(f) \in \mathbb{R} \cup \{\pm\infty\}$) such that $\{\mathrm{Re}\, s > D(f)\}$ is the *maximal* open right half-plane (of the form $\{\mathrm{Re}\, s > \alpha\}$, for some $\alpha \in \mathbb{R} \cup \{\pm\infty\}$) on which the Lebesgue integral initially defining f in (2.5) is convergent (or, equivalently, is absolutely convergent), with μ replaced by $|\mu|$, the total variation measure of μ. (Recall that $|\mu| = \mu$ if μ is positive.) In short, $D(f)$ is called the *abscissa of convergence* of f. Furthermore, the aforementioned maximal right half-plane is denoted by $\Pi(f)$ and is called the *half-plane of* (absolute or Lebesgue) *convergence* of (the Dirichlet-type integral) f. It is shown in [41, §2.1]

---

[8] By using the principle of analytic continuation, it is easy to check that $D_{\mathrm{hol}}(f)$ and $\mathcal{H}(f)$ are well defined; see [41, §2.1].

[9] This is the case of the classic (generalized) Dirichlet series and integrals [55, 58], the classic arithmetic zeta functions (see, e.g., [52, App. A] and [25, Apps. B, C & E]), as well as of the geometric zeta functions of fractal strings studied in [50–52] and of all the fractal zeta functions considered in this paper and in [41–47].

[10] Such functions f are called *tamed* DTIs in [41–47]; see esp. [41, App. A] for a development of their general theory.

that under mild hypotheses (which are always satisfied in our setting), $D(f)$ is well defined and (with the notation of (2.5) just above) we have, equivalently[11]:

$$D(f) = \inf\left\{\alpha \in \mathbb{R} : \int_E \varphi(x)^\alpha \mathrm{d}|\mu|(x) < \infty\right\}, \tag{2.7}$$

where (as above) $|\mu|$ is the total variation (local) measure of μ. Under the stated conditions on f, we have $\Pi(f) \subseteq \mathcal{H}(f)$; that is, $-\infty \leq D_{\mathrm{hol}}(f) \leq D(f) \leq +\infty$.

Note that the distance zeta function ζ_A, defined by (2.1), is a tamed DTI of the form (2.5), with $E := A_\delta$, $\varphi(x) := d(x, A)$ and $\mathrm{d}\mu(x) := d(x, A)^{-N}\mathrm{d}x$. Furthermore, we can clearly take $C := \delta$ in (2.6).

The following key result describes some of the basic properties of distance zeta functions.

Theorem 2.2 ([41, 42]) *Let A be an arbitrary bounded subset of $\mathbb{R}^N$ and let δ be a fixed positive real number. Then*:

(*a*) *The distance zeta function ζ_A is holomorphic on* $\{\mathrm{Re}\, s > \overline{\dim}_B A\}$. *Moreover,* $\Pi(\zeta_A) = \{\mathrm{Re}\, s > \overline{\dim}_B A\}$*; that is,*

$$D(\zeta_A) = \overline{\dim}_B A. \tag{2.8}$$

(*b*) *If the box* (*or Minkowski*) *dimension $D := \dim_B A$ exists, $D < N$ and $\mathcal{M}_*^D(A) > 0$, then $\zeta_A(s) \to +\infty$ as $s \in \mathbb{R}$ converges to D from the right. In particular,* $\mathcal{H}(\zeta_A) = \Pi(\zeta_A) = \{\mathrm{Re}\, s > \dim_B A\}$*; that is,*

$$D_{\mathrm{hol}}(\zeta_A) = D(\zeta_A) = \dim_B A. \tag{2.9}$$

Remark 3

(*a*) It would be of interest to construct (if possible) a class of nontrivial bounded subsets A of $\mathbb{R}^N$ such that $D_{\mathrm{hol}}(\zeta_A) < D(\zeta_A)$. A trivial example is given by $A = [0, 1]$, since then $D_{\mathrm{hol}}(\zeta_A) = 0$ and $D(\zeta_A) = 1$.

(*b*) The analog of Theorem 2.2 holds for the tube zeta function $\tilde{\zeta}_A$ (to be introduced in Definition 2.4 below), except for the fact that in part (*b*), one no longer needs to assume that $D < N$.

Given a bounded set A, it is of interest to know the corresponding poles of the associated distance zeta function ζ_A, meromorphically extended (if possible) to a neighborhood of the critical line $\{\mathrm{Re}\, s = D(\zeta_A)\}$. Following the terminology of [52], these poles are called the *complex dimensions* of A and we denote the resulting

[11] Let $D := \overline{\dim}_B A$, for brevity. In light of Theorem 2.2, for this alternative definition of $D(\zeta_A)$ (or of $D(\tilde{\zeta}_A)$), with $A \subseteq \mathbb{R}^N$ bounded (as in the present situation), it would suffice to restrict oneself to $\alpha \geq 0$ in the right-hand side of (2.7); this follows since $D(\zeta_A) = \overline{\dim}_B A \geq 0$ and (if $D < N$), $D(\zeta_A) = D(\tilde{\zeta}_A)$. Here, $\tilde{\zeta}_A$ stands for the tube zeta function of A, defined by Eq. (2.12).

set of complex dimensions by $\mathcal{P}(\zeta_A)$.[12] We pay particular attention to the set of complex dimensions of A located on the critical line $\{\operatorname{Re} s = D(\zeta_A)\}$, which we call the set of *principal complex dimensions* of A and denote by $\dim_{PC} A$.

For example, it is well known that for the ternary Cantor set $C^{(1/3)}$, $\dim_B C^{(1/3)} = \log_3 2$ and, moreover (see [52, §1.2.2 and §2.3.1]), with $\mathbb{i} := \sqrt{-1}$,

$$\dim_{PC} C^{(1/3)} := \log_3 2 + \frac{2\pi}{\log 3}\,\mathbb{i}\mathbb{Z}.$$

The following result provides an interesting connection between the residue of the distance zeta function of a fractal set at $D := \dim_B A$ and its Minkowski contents.

Theorem 2.3 ([41, 42]) *Assume that A is a bounded subset of $\mathbb{R}^N$ which is nondegenerate (that is, $0 < \mathcal{M}_*^D(A) \le \mathcal{M}^{*D}(A) < \infty$ and, in particular, $\dim_B A = D$), and $D < N$. If the distance zeta function $\zeta_A(\,\cdot\,, A_\delta) := \zeta_A$, initially defined by (2.1), can be meromorphically extended*[13] *to a neighborhood of $s = D$, then D is necessarily a simple pole of $\zeta_A(\,\cdot\,, A_\delta)$, and*

$$(N-D)\,\mathcal{M}_*^D(A) \le \operatorname{res}(\zeta_A(\,\cdot\,, A_\delta), D) \le (N-D)\,\mathcal{M}^{*D}(A). \tag{2.10}$$

Furthermore, the value of $\operatorname{res}(\zeta_A(\,\cdot\,, A_\delta), D)$ *does not depend on $\delta > 0$. In particular, if A is Minkowski measurable, then*

$$\operatorname{res}(\zeta_A(\,\cdot\,, A_\delta), D) = (N-D)\,\mathcal{M}^D(A). \tag{2.11}$$

The distance zeta function defined by (2.1) is closely related to the tube zeta function of a fractal set, which, in turn, is defined via the tube function $t \mapsto |A_t|$, for $t > 0$, of the fractal set A, as we now explain.

Definition 2.4 ([41, 42]) Let δ be a fixed positive number, and let A be a bounded subset of $\mathbb{R}^N$. Then, the *tube zeta function* of A, denoted by $\tilde{\zeta}_A$, is defined (for all $s \in \mathbb{C}$ with $\operatorname{Re} s$ sufficiently large) by

$$\tilde{\zeta}_A(s) := \int_0^\delta t^{s-N-1} |A_t|\,\mathrm{d}t. \tag{2.12}$$

[12] Strictly speaking, one should talk about the set $\mathcal{P}(\zeta_A, U)$ of *visible complex dimensions* relative to a domain $U \subseteq \mathbb{C}$ to which ζ_A can be meromorphically extended; see [41–44] (along with [52]). In the examples described in this paper, we have $U := \mathbb{C}$.

[13] The existence and construction of meromorphic extensions of fractal zeta functions is discussed in Sect. 3. It is studied in a variety of situations in [41–44, 46].

For any fixed positive real number $\delta > 0$, the distance and tube zeta functions associated with a given fractal set A are connected as follows[14]:

$$\zeta_A(s, A_\delta) = \delta^{s-N}|A_\delta| + (N-s)\tilde{\zeta}_A(s,\delta), \tag{2.13}$$

for $\mathrm{Re}\, s > \overline{\dim}_B A$[15]; see [41, 42].[16] Using this result, it is easy to obtain the analog of Theorem 2.2 for $\tilde{\zeta}_A$ (as was stated in Remark 3(*b*) above) and to reformulate Theorem 2.3 in terms of the tube zeta functions. In particular, we conclude that the residue of the tube zeta function of a fractal set, computed at $s = D$, is equal to its Minkowski content, provided the set is Minkowski measurable.

Theorem 2.5 ([41, 42]) *Assume that A is a nondegenerate bounded subset of $\mathbb{R}^N$ (so that $D := \dim_B A$ exists), and there exists a meromorphic extension of $\tilde{\zeta}_A$ to a neighborhood of D. Then, D is a simple pole of $\tilde{\zeta}_A$, and for any positive δ, $\mathrm{res}(\tilde{\zeta}_A, D)$ is independent of δ. Furthermore, we have*

$$\mathcal{M}_*^D(A) \le \mathrm{res}(\tilde{\zeta}_A, D) \le \mathcal{M}^{*D}(A). \tag{2.14}$$

In particular, if A is Minkowski measurable, then

$$\mathrm{res}(\tilde{\zeta}_A, D) = \mathcal{M}^D(A). \tag{2.15}$$

A class of fractal sets A for which we have strict inequalities in (2.14) (and hence also in (2.10) of Theorem 2.3 above) is constructed in Theorem 3.2; see (3.8).

3 Meromorphic Extensions of Fractal Zeta Functions

Since the definition of the set of principal complex dimensions $\dim_{PC} A$ of A requires the existence of a suitable meromorphic extension of the distance zeta function ζ_A, it is natural to study this issue in more detail. For simplicity, we formulate the results of this section for $\tilde{\zeta}_A$, but we note that the analogs of Theorems 3.1 and 3.2 also hold for ζ_A, provided $D < N$; see [41, §2.3.3] or [43].

Theorem 3.1 (Minkowski measurable case, [41, 43]) *Let A be a bounded subset of $\mathbb{R}^N$ such that there exist $\alpha > 0$, $\mathcal{M} \in (0, +\infty)$ and $D \ge 0$ satisfying*

$$|A_t| = t^{N-D}\left(\mathcal{M} + O(t^\alpha)\right) \quad \text{as} \quad t \to 0^+. \tag{3.1}$$

[14]We write here $\zeta_A(\,\cdot\,, A_\delta) := \zeta_A$ and $\tilde{\zeta}_A(\,\cdot\,,\delta) := \tilde{\zeta}_A$, for emphasis.

[15]In light of the principle of analytic continuation, one deduces that identity (2.13) continues to hold whenever one (and hence, both) of the fractal zeta functions ζ_A and $\tilde{\zeta}_A$ is meromorphic on a given domain $U \subseteq \mathbb{C}$.

[16]The case when $D = N$ in Theorem 2.5 must be treated separately.

Then, $\dim_B A$ *exists and* $\dim_B A = D$*. Furthermore, A is Minkowski measurable with Minkowski content* $\mathcal{M}^D(A) = \mathcal{M}$*. Moreover, the tube zeta function* $\tilde{\zeta}_A$ *has for abscissa of convergence* $D(\tilde{\zeta}_A) = \dim_B A = D$ *and possesses a (necessarily unique) meromorphic continuation (still denoted by* $\tilde{\zeta}_A$*) to (at least) the open right half-plane* $\{\mathrm{Re}\, s > D - \alpha\}$*. The only pole of* $\tilde{\zeta}_A$ *in this half-plane is* $s = D$*; it is simple and, moreover,* $\mathrm{res}(\tilde{\zeta}_A, D) = \mathcal{M}$*.*

Next, we deal with a useful class of Minkowski nonmeasurable sets. Before stating Theorem 3.2, let us first introduce some notation. Given a locally integrable T-periodic function $G : \mathbb{R} \to \mathbb{R}$, with $T > 0$, we denote by G_0 its truncation to $[0, T]$, while the Fourier transform of G_0 is denoted by $\hat{G}_0$: for all $t \in \mathbb{R}$,

$$\hat{G}_0(t) := \int_{-\infty}^{+\infty} \mathrm{e}^{-2\pi \mathrm{i} t\tau} G_0(\tau)\,\mathrm{d}\tau = \int_0^T \mathrm{e}^{-2\pi \mathrm{i} t\tau} G(\tau)\,\mathrm{d}\tau. \tag{3.2}$$

Theorem 3.2 (Minkowski nonmeasurable case, [41, 43]) *Let A be a bounded subset of* $\mathbb{R}^N$ *such that there exist* $D \geq 0$*,* $\alpha > 0$*, and* $G : \mathbb{R} \to (0, +\infty)$ *a nonconstant periodic function with period* $T > 0$*, satisfying*

$$|A_t| = t^{N-D}\left(G(\log t^{-1}) + O(t^\alpha)\right) \quad \text{as} \quad t \to 0^+. \tag{3.3}$$

Then G is continuous, $\dim_B A$ *exists and* $\dim_B A = D$*. Furthermore, A is Minkowski nondegenerate, with upper and lower Minkowski contents respectively given by*

$$\mathcal{M}_*^D(A) = \min G, \quad \mathcal{M}^{*D}(A) = \max G. \tag{3.4}$$

Moreover, the tube zeta function $\tilde{\zeta}_A$ *has for abscissa of convergence* $D(\tilde{\zeta}_A) = D$ *and possesses a (necessarily unique) meromorphic extension (still denoted by* $\tilde{\zeta}_A$*) to (at least) the half-plane* $\{\mathrm{Re}\, s > D - \alpha\}$*.*

In addition, the set of principal complex dimensions of A is given by

$$\dim_{PC} A = \left\{ s_k = D + \frac{2\pi}{T}\mathrm{i}k : \hat{G}_0\Big(\frac{k}{T}\Big) \neq 0,\ k \in \mathbb{Z} \right\} \tag{3.5}$$

(see (3.2)) and there are no other complex dimensions in $\{\mathrm{Re}\, s > D - \alpha\}$*; they are all simple, and the residue at each* $s_k \in \dim_{PC} A$*, with* $k \in \mathbb{Z}$*, is given by*

$$\mathrm{res}(\tilde{\zeta}_A, s_k) = \frac{1}{T}\hat{G}_0\Big(\frac{k}{T}\Big). \tag{3.6}$$

If $s_k \in \dim_{PC} A$*, then* $s_{-k} \in \dim_{PC} A$ *(in agreement with the 'reality principle'), and* $|\,\mathrm{res}(\tilde{\zeta}_A, s_k)| \leq \frac{1}{T}\int_0^T G(\tau)\,\mathrm{d}\tau$*; furthermore,* $\lim_{k \to \pm\infty} \mathrm{res}(\tilde{\zeta}_A, s_k) = 0$*.*

Moreover, the set of principal complex dimensions of A contains $s_0 = D$*, and*

$$\mathrm{res}(\tilde{\zeta}_A, D) = \frac{1}{T}\int_0^T G(\tau)\,\mathrm{d}\tau. \tag{3.7}$$

In particular, A is not Minkowski measurable and

$$\mathcal{M}_*^D(A) < \operatorname{res}(\tilde{\zeta}_A, D) < \mathcal{M}^{*D}(A). \tag{3.8}$$

Example 1 (a-strings) The compact set $A := \{j^{-a} : j \in \mathbb{N}\} \cup \{0\}$, where $a > 0$, is Minkowski measurable and

$$\mathcal{M}^D(A) = \frac{2^{1-D}}{1-D} a^D, \quad D := \dim_B A = \frac{1}{1+a}. \tag{3.9}$$

(See [22, Exple. 5.1 and App. C].) The associated fractal string $\mathcal{L} = (\ell_j)_{j \geq 1}$, defined by $\ell_j = j^{-a} - (j+1)^{-a}$ for all $j \geq 1$ (or, equivalently, by $\Omega := [0,1] \setminus A \subset \mathbb{R}$, so that $\partial\Omega = A$), is called the *a-string*; see [14, 22–24, 38, 39] and [52, §6.5.1]. In light of (2.11) and (2.15), we then know that $\operatorname{res}(\zeta_A(\,\cdot\,, A_\delta), D) = (1-D)\mathcal{M}^D(A)$ and $\operatorname{res}(\tilde{\zeta}_A, D) = \mathcal{M}^D(A)$.

Example 2 (fractal nests) Let $a > 0$ and let A be the countable union of concentric circles in $\mathbb{R}^2$, centered at the origin and of radii $r = k^{-a}$, where $k \in \mathbb{N}$. According to the terminology introduced in [41–44], A is called the *fractal nest of inner type* generated by the a-string from the preceding example. Then, using the distance zeta function of A it is possible to show that

$$D := \overline{\dim}_B A = \max\left\{1, \frac{2}{1+a}\right\}. \tag{3.10}$$

(See [41, Chap. 3] and [42–44].) The set A is closely related to the planar spiral Γ defined in polar coordinates by $r = \theta^{-a}$, $\theta \geq \theta_0$, where $\theta_0 > 0$, and the value of $\dim_B \Gamma$ is the same as for A; see [63]. We mention in passing that for $a \neq 1$, the fractal nest A (as well as the corresponding spiral Γ) is Minkowski measurable and for every $a \in (0,1)$, the value of its Minkowski content is independent of θ_0 and given by

$$\mathcal{M}^D(A) = \pi(2/a)^{2a/(1+a)} \frac{1+a}{1-a}. \tag{3.11}$$

Using (3.11), along with Eq. (2.11) from Theorem 2.3, we conclude that the residue of the distance zeta function ζ_A, computed at $s = D$, is given by

$$\operatorname{res}(\zeta_A, D) = \pi(2/a)^{2a/(1+a)} \frac{2a}{1-a}, \tag{3.12}$$

provided $a \in (0,1)$. For $a = 1$, we have $\mathcal{M}^1(A) = \mathcal{M}^1(\Gamma) = +\infty$. These and related results are useful in the study of fractal properties of spiral trajectories of planar vector fields; see, e.g., [69].

More generally, if we consider the fractal nest A_N defined as the countable union of concentric spheres in $\mathbb{R}^N$, centered at the origin and of radii $r = k^{-\alpha}$, where $k \in \mathbb{N}$, then using the distance zeta function ζ_{A_N}, it can be shown (see [41, §3.4] and [42–44]) that

$$\overline{\dim}_B A_N = \max\left\{N-1, \frac{N}{1+a}\right\}. \tag{3.13}$$

Note that for $N = 1$ and $N = 2$, we recover the box dimension of the a-string and of the fractal nest, respectively; see [42–44] and Eqs. (3.9)–(3.10) above.

In the following result, we provide the distance zeta function of the Sierpiński carpet and the corresponding principal complex dimensions. It is well known that the Sierpiński carpet is not Minkowski measurable. See, e.g., [52], as well as [17] for explicit values of its upper and lower Minkowski contents. A similar result can be obtained for the Sierpiński gasket (and its higher-dimensional analogs); see [41, §3.2.2] and [42–44].

Proposition 3.3 (Distance zeta function of the Sierpiński carpet) *Let A be the Sierpiński carpet in $\mathbb{R}^2$, constructed in the usual way inside the unit square. Let δ be a fixed positive real number. We assume without loss of generality that $\delta > 1/6$ (so that for this choice of δ, A_δ coincides with the δ-neighborhood of the unit square* $[0,1]^2$). *Then, for all $s \in \mathbb{C}$, the distance zeta function ζ_A of the Sierpiński carpet is given by*

$$\zeta_A(s) = \frac{8}{2^s s(s-1)(3^s-8)} + 2\pi\frac{\delta^s}{s} + 4\frac{\delta^{s-1}}{s-1}, \tag{3.14}$$

which is meromorphic on the whole complex plane. In particular, the set of complex dimensions and of principal complex dimensions of the Sierpiński carpet are given, respectively, by

$$\mathcal{P}(\zeta_A) = \{0,1\} \cup \dim_{PC} A, \quad \dim_{PC} A = \log_3 8 + \frac{2\pi}{\log 3}\mathbb{i}\mathbb{Z}. \tag{3.15}$$

Furthermore, each of the complex dimensions (*i.e., each of the poles of* ζ_A) *is simple. Moreover, the residues of the distance zeta function ζ_A computed at the principal poles $s_k := \log_3 8 + \frac{2\pi}{\log 3}k\mathbb{i}$, with $k \in \mathbb{Z}$, are given by*

$$\operatorname{res}(\zeta_A, s_k) = \frac{2^{-s_k}}{(\log 3)s_k(s_k-1)}. \tag{3.16}$$

Finally, the approximate values of the lower and upper D-dimensional Minkowski contents are given by $\mathcal{M}_*^D(A) \approx 1.350670$ and $\mathcal{M}^{*D}(A) \approx 1.355617$. (*The precise values can be found in* [17].)

Sketch of the proof In order to compute the distance zeta function

$$\zeta_A(s) := \int_{A_\delta} d((x,y),A)^{s-2} \mathrm{d}x\,\mathrm{d}y$$

of the Sierpiński carpet A, we first have to calculate

$$\zeta_{A_k}(s,\Omega_k) := \int_{\Omega_k} d((x,y),A_k)\,\mathrm{d}x\,\mathrm{d}y, \tag{3.17}$$

where Ω_k is a square of the k-th generation (of side lengths $a_k = 3^{-k}$) and A_k is its boundary. (Here, we deal in fact with 'relative distance zeta functions', which are discussed in Remark 4 just below; see (6.2) and [41–44].) This can be easily done by splitting Ω_k into the disjoint union of eight congruent right-angle triangles, and we obtain after a short computation that $\zeta_{A_k}(s,\Omega_k) = 8\cdot 2^{-s}a_k^s s^{-1}(s-1)^{-1}$. Since the k-th generation consists of 8^{k-1} squares congruent to Ω_k, we deduce that

$$\zeta_A(s,[0,1]^2) = \sum_{k=1}^{\infty} 8^{k-1}\zeta_{A_k}(s,\Omega_k) = \frac{8}{2^s s(s-1)(3^s-8)}, \tag{3.18}$$

for $\operatorname{Re} s > \log_3 8$. The last expression in (3.18) is meromorphic in all of $\mathbb{C}$. Hence, upon analytic continuation, $\zeta_A(s,[0,1]^2)$ is given by that expression for all $s\in\mathbb{C}$. Note that the value of $\zeta_A(s,[0,1]^2)$ is precisely equal to the first term on the right-hand side of (3.14). The remaining two terms are obtained by considering $\zeta_A(s,A_\delta\setminus[0,1]^2)$, which can be easily reduced to considering a disk $B_\delta(0)$ of radius δ with respect to its origin $0\in\mathbb{R}^2$, and two rectangles that are congruent to $\Omega_0 := (0,1)\times(-\delta,\delta)$ with respect to its middle section $A_0 := (0,1)\times\{0\}$. □

Remark 4 Equation (3.17) is a very special case of the zeta function of a *relative fractal drum* (A,Ω) in $\mathbb{R}^N$, a notion which will be briefly discussed in Sect. 6 and is the object of [44] and [41, Chap. 4]; see the first equality in Eq. (6.2) below.

4 Transcendentally Quasiperiodic Sets

In this section, we define a class of *quasiperiodic fractal sets*. The simplest of such sets has two incommensurable periods. Moreover, using suitable generalized Cantor sets, it is possible to ensure that the quotient of their periods be a transcendental real number. Our construction of such sets is based on a class of generalized Cantor sets with two parameters, which we now introduce.

Definition 4.1 ([41, 42]) The generalized Cantor sets $C^{(m,a)}$ are determined by an integer $m\geq 2$ and a real number $a\in(0,1/m)$. In the first step of the analog of Cantor's construction, we start with m equidistant, closed intervals in $[0,1]$ of length a, with $m-1$ 'holes', each of length $(1-ma)/(m-1)$. In the second step, we continue by scaling by the factor a each of the m intervals of length a; and so on, ad infinitum. The (two-parameter) *generalized Cantor set* $C^{(m,a)}$ is then defined as the intersection of the decreasing sequence of compact sets constructed in this way. It is easy to check that $C^{(m,a)}$ is a perfect, uncountable compact subset of $\mathbb{R}$; furthermore, $C^{(m,a)}$ is also self-similar. For $m = 2$, the sets $C^{(m,a)}$ are denoted by $C^{(a)}$. The classic ternary Cantor set is obtained as $C^{(2,1/3)}$. In order to avoid any possible confusion, we note that the generalized Cantor sets introduced here are different from the generalized Cantor strings introduced and studied in [52, Chap. 10], as well as used in a key manner in [52, Chap. 11].

We collect some of the basic properties of generalized Cantor sets in the following proposition.

Proposition 4.2 (Generalized Cantor sets, [41, 42]) *If $A := C^{(m,a)}$ is the generalized Cantor set introduced in Definition 4.1, where m is an integer larger than* 1*, and $a \in (0, 1/m)$, then*

$$D := \dim_B A = D(\zeta_A) = \log_{1/a} m. \tag{4.1}$$

Furthermore, the tube formula associated with A is given by

$$|A_t| = t^{1-D} G(\log t^{-1}) \quad \textit{for all} \quad t \in (0, t_0), \tag{4.2}$$

where t_0 is a suitable positive constant and $G = G(\tau)$ is a continuous, positive and nonconstant periodic function, with minimal period $T = \log(1/a)$.

Moreover, A is Minkowski nondegenerate and Minkowski nonmeasurable; that is, $0 < \mathcal{M}_*^D(A) < \mathcal{M}^{*D}(A) < \infty$.[17]

Finally, the distance zeta function of A admits a meromorphic continuation to all of $\mathbb{C}$ and the set of principal complex dimensions of A is given by

$$\dim_{PC} A = D + \frac{2\pi}{T}\mathbb{i}\mathbb{Z}. \tag{4.3}$$

Besides $(\dim_{PC} A) \cup \{0\}$*, there are no other poles, and all of the poles of ζ_A are simple. In particular,* $\mathcal{P}(\zeta_A) = (D + \frac{2\pi}{T}\mathbb{i}\mathbb{Z}) \cup \{0\}$.

The definition of quasiperiodic sets is based on the following notion of quasiperiodic functions, which will be useful for our purposes.[18]

[17] The periodic function $G = G(\tau)$, as well as the values of $\mathcal{M}_*^D(A)$ and $\mathcal{M}^{*D}(A)$, can be explicitly computed; see [41, §3.1.1].

[18] We note that Definition 4.3, although rather close to the one provided in [64], is very different from the usual definition of Bohr-type quasiperiodic functions.

Definition 4.3 ([41, 42]) We say that a function $G = G(\tau) : \mathbb{R} \to \mathbb{R}$ is *transcendentally n-quasiperiodic*, with $n \geq 2$, if it is of the form $G(\tau) = H(\tau, \ldots, \tau)$, where $H : \mathbb{R}^n \to \mathbb{R}$ is a function that is nonconstant and T_k-periodic in its k-th component, for each $k = 1, \ldots, n$, and the periods $T_1, \ldots, T_n$ are algebraically (and hence, rationally) independent.[19] The positive numbers T_i ($i = 1, \ldots, n$) are called the *quasiperiods* of G. If, instead, the set of quasiperiods $\{T_1, \ldots, T_n\}$ is rationally independent and algebraically dependent, we say that G *is algebraically n-quasiperiodic*.

Definition 4.4 ([41, 42]) Given a bounded subset A of $\mathbb{R}^N$, we say that a function $G : \mathbb{R} \to \mathbb{R}$ *is associated with the set A* (or *corresponds to A*) if it is nonnegative and A has the following tube formula:

$$|A_t| = t^{N-D}(G(\log(1/t)) + o(1)) \text{ as } t \to 0^+, \tag{4.4}$$

with $0 < \liminf_{\tau\to+\infty} G(\tau) \leq \limsup_{\tau\to+\infty} G(\tau) < \infty$. In addition, we say that A is a *transcendentally* (resp., *algebraically*) *n-quasiperiodic set* if the function $G = G(\tau)$ is transcendentally (resp., algebraically) n-quasiperiodic. The smallest possible value of n is called the *order of quasiperiodicity* of A.

The following result, which has a variety of generalizations as will be briefly explained below, provides a construction of transcendentally 2-quasiperiodic fractal sets. Its proof is based on the classical Gel'fond–Schneider theorem (as described in [10]) from transcendental number theory.

Theorem 4.5 ([41, 42]) *Let $C^{(m_1,a_1)}$ and $\dim_B C^{(m_2,a_2)}$ be two generalized Cantor sets such that their box dimensions coincide and are equal to $D \in (0,1)$. Assume that I_1 and I_2 are two unit closed intervals of $\mathbb{R}$, with disjoint interiors, and define $A_1 := (\min I_1) + C^{(m_1,a_1)} \subset I_1$ and $A_2 := (\min I_2) + C^{(m_2,a_2)} \subset I_2$. Let $\{p_1, p_2, \ldots, p_k\}$ be the set of all distinct prime factors of m_1 and m_2, and write*

$$m_1 = p_1^{\alpha_1} p_2^{\alpha_2} \ldots p_k^{\alpha_k}, \quad m_2 = p_1^{\beta_1} p_2^{\beta_2} \ldots p_k^{\beta_k},$$

where $\alpha_i, \beta_i \in \mathbb{N} \cup \{0\}$ for $i = 1, \ldots, k$. If the exponent vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ of, respectively, m_1 and m_2, defined by

$$\mathbf{e}_1 := (\alpha_1, \alpha_2, \ldots, \alpha_k) \quad \text{and} \quad \mathbf{e}_2 := (\beta_1, \beta_2, \ldots, \beta_k),$$

are linearly independent over the field of rational numbers, then the compact set $A := A_1 \cup A_2 \subset \mathbb{R}$ is transcendentally 2-quasiperiodic.

Moreover, the distance zeta function ζ_A can be meromorphically extended to the whole complex plane, and we have that $D(\zeta_A) = D$. The set $\dim_{PC} A$ of principal

[19] That is, linearly independent over the field of algebraic numbers.

complex dimensions of A is given by

$$\dim_{PC} A = D + \left(\frac{2\pi}{T_1}\mathbb{Z} \cup \frac{2\pi}{T_2}\mathbb{Z}\right)\mathbb{i}. \tag{4.5}$$

Besides $(\dim_{PC} A) \cup \{0\}$, *there are no other poles of the distance zeta function* ζ_A *and they are all simple. In particular,*

$$\mathcal{P}(\zeta_A) = \left(D + \left(\frac{2\pi}{T_1}\mathbb{Z} \cup \frac{2\pi}{T_2}\mathbb{Z}\right)\mathbb{i}\right) \cup \{0\}. \tag{4.6}$$

Remark 5 This result can be considerably extended by using Baker's theorem [1, Thm. 2.1] which, in turn, is a far-reaching extension of the aforementioned Gel'fond–Schneider's theorem. Indeed, for any fixed integer $n \geq 2$, using Baker's theorem and n generalized Cantor sets, an explicit construction of a class of transcendentally n-quasiperiodic fractal sets is given in [42] and [41, §3.1]. In [43, 44] and [41, Chap. 4], we even construct a set which is transcendentally ∞-quasiperiodic; see Sect. 5.

5 Maximally Hyperfractal ∞-Quasiperiodic Sets

It is possible to construct a bounded subset A of the real line, such that the corresponding distance zeta function ζ_A has for abscissa of (Lebesgue, i.e., absolute) convergence $D(\zeta_A)$ any prescribed real number $D \in (0, 1)$ and A is *maximally hyperfractal*; that is, any point on the critical line $\{\operatorname{Re} s = D\}$ is a nonremovable singularity of the corresponding distance zeta function ζ_A. In particular, there is no meromorphic continuation of ζ_A to any open and connected neighborhood of the critical line (and, moreover, not even to any open and connected neighborhood of an arbitrary point on the critical line). Furthermore, it is possible to construct a maximally hyperfractal set which is ∞-transcendentally quasiperiodic as well. A construction of such sets is described in detail in [41, Chap. 4] or in [44]. In the sequel, we provide a rough sketch of this construction.

The set $A \subset \mathbb{R}$ which is a maximal hyperfractal and ∞-transcendentally quasiperiodic set, can be constructed as the nonincreasing sequence

$$A = A_{\mathcal{L}} = \left\{a_k := \sum_{j=k}^{\infty} \ell_j : k \in \mathbb{N}\right\} \tag{5.1}$$

of positive real numbers a_k converging to zero as $k \to \infty$, generated by a suitable bounded fractal string $\mathcal{L} = (\ell_j)_{j\geq 1}$. Roughly speaking, the fractal string $\mathcal{L}$ is obtained as a (suitably defined) *union of an infinite sequence of bounded fractal strings* $\mathcal{L}_k := (\ell_{kj})_{j\geq 1}$, corresponding to generalized Cantor sets of the form

$c_k \cdot C^{(m_k,a_k)}$, for $k \in \mathbb{N}$, with carefully chosen values of the parameters m_k and a_k appearing in Definition 4.1, and where $(c_k)_{k\geq 1}$ is an appropriate summable sequence of positive real numbers.

More precisely, the union $\mathcal{L} := \bigsqcup_{k=1}^{\infty} \mathcal{L}_k$ of the sequence of bounded fractal strings $\mathcal{L}_k$ is defined as the set-theoretic union of the elements of the strings, but by definition, each of its elements has for multiplicity the sum of the corresponding multiplicities from all of the fractal strings $\mathcal{L}_k$ to which belongs the element in question. Note that the multiplicity of an element of $\mathcal{L}$ is well defined since this element must belong to at most finitely many bounded fractal strings $\mathcal{L}_k$, which follows from the fact that the sequence c_k converges to 0 as $k \to \infty$. Moreover, we must assume that $\sum_{k=1}^{\infty} c_k < \infty$, so that the string $\mathcal{L}$ be bounded (i.e., $\sum_{j=1}^{\infty} \ell_j < \infty$). We can also ensure that for each positive integer k, the corresponding upper box dimension of $\mathcal{L}_k$ (that is, of the set $A_{\mathcal{L}_k}$) be equal to a fixed value of $D \in (0, 1)$, prescribed in advance. (Note that the set $A_{\mathcal{L}}$ is distinct from $\cup_{k=1}^{\infty} A_{\mathcal{L}_k}$.)

Recall that the oscillatory period of $\mathcal{L}_k$ (in the sense of [52]), which is defined by $\mathbf{p}_k := \frac{2\pi}{\log(1/a_k)}$, provides valuable information about the density of the set of principal complex dimensions of $\mathcal{L}_k$ on the critical line $\{\mathrm{Re}\, s = D\}$. More precisely, by choosing the coefficient $a_k \in (0, 1/m_k)$ so that $a_k \to 0$ as $k \to \infty$, we see that for the set of principal complex dimensions of the generalized Cantor string $\mathcal{L}_k$ (i.e., the set of the principal poles of $\zeta_{\mathcal{L}_k}$), $\dim_{PC} \mathcal{L}_k = \dim_{PC} C^{(m_k,a_k)} = D + \mathbf{p}_k \mathbb{i}\mathbb{Z}$, becomes denser and denser on the critical line, as $k \to \infty$, since then the oscillatory period $\mathbf{p}_k$ tends to zero. Therefore, the distance zeta function of the fractal string $\mathcal{L} := \sqcup_{k=1}^{\infty} \mathcal{L}_k$ will have $D + \left(\bigcup_{k=1}^{\infty} \mathbf{p}_k \mathbb{Z}\right)\mathbb{i}$ as a set of singularities, which is densely packed on the critical line $\{\mathrm{Re}\, s = D\} = D + \mathbb{R}\mathbb{i}$, since the set $\cup_{k=1}^{\infty} \mathbf{p}_k \mathbb{Z}$ is clearly dense in $\mathbb{R}$. In conclusion, the whole critical line $\{\mathrm{Re}\, s = D\}$ consists of nonremovable singularities of $\zeta_{\mathcal{L}}$,[20] which by definition means that the fractal string $\mathcal{L}$ is maximally hyperfractal. Hence, the corresponding set $A := A_{\mathcal{L}}$ is also maximally hyperfractal.

Since the coefficients a_k, appearing in the definition of the generalized Cantor set (see Definition 4.1), have been chosen above so that $a_k \to 0$ as $k \to \infty$, it is clear that $m_k \to \infty$, because $D = \dim_B C^{(m_k,a_k)} = \frac{m_k}{\log(1/a_k)}$, where $D \in (0, 1)$ is given in advance and independent of k. This enables us to use our result mentioned in Remark 5, obtained by means of Baker's theorem from transcendental number theory [1], in order to ensure that the sequence of quasiperiods $T_k := \log(1/a_k)$, $k \in \mathbb{N}$, is algebraically independent (that is, any finite subset of this set of quasiperiods is linearly independent over the field of algebraic real numbers).[21] According to

[20] In light of the discussion surrounding Eq. (2.4) above, the same is true if $\zeta_{\mathcal{L}}$ is replaced by $\zeta_{A_{\mathcal{L}}}$ or, more generally, by the *relative distance zeta function* $\zeta_A(\,\cdot\,, \Omega)$ defined by $\zeta_A(s, \Omega) := \int_{\Omega} d(x, A)^{s-N} \mathrm{d}x$ (see Sect. 6 and [44] or [41, Chap. 4]), where $A = \partial\Omega$ is the boundary of any geometric realization of $\mathcal{L}$ by a bounded open subset Ω of $\mathbb{R}$.

[21] The algebraic independence of the set of quasiperiods $\{T_k : k \geq 1\}$, with $k \geq 1$, can be deduced (using the aforementioned Baker's theorem, [1]) if we assume, in addition, that the sequence $(\mathbf{e}_k)_{k\geq 1}$ (suitably redefined), corresponding to the sequence $(m_k)_{k\geq 1}$, is rationally independent.

Definition 4.3, this means that $\mathcal{L}$ is ∞-transcendentally quasiperiodic, and so is the corresponding bounded subset $A := A_{\mathcal{L}}$ of the real line.

As we see from the above rough description, the nature of a subset $A := A_{\mathcal{L}}$ of the real line which is maximally hyperfractal and ∞-transcendentally quasiperiodic, is in general extremely complex, although it is, in fact, 'just' defined in terms of a nonincreasing sequence of positive real numbers converging to zero.

In closing this discussion, we mention that this construction (as well as Theorem 4.5 and its generalization mentioned in Remark 5), extends to any $N \geq 2$, by letting $B := A \times [0,1]^{N-1} \subset \mathbb{R}^N$; see [41, 42, 44].

6 Fractal Zeta Functions of Relative Fractal Drums

In this section, we survey some of the definitions and results from [41, Chap. 4]; see also [44]. Let A be a (possibly unbounded) subset of $\mathbb{R}^N$ and let Ω be a (possibly unbounded) Borel subset of $\mathbb{R}^N$ of finite N-dimensional Lebesgue measure. We say that the ordered pair (A, Ω) is a *relative fractal drum* (or RFD, in short) if there exists a positive real number δ such that $\Omega \subseteq A_\delta$. It is easy to see that for every $\delta > 0$, any bounded subset A can be identified with the relative fractal drum (A, A_δ). Furthermore, any bounded fractal string $\mathcal{L} = (\ell_j)_{j=1}^{\infty}$ can be identified with the relative fractal drum $(\cup_{j=1}^{\infty} \partial I_j, \cup_{j=1}^{\infty} I_j)$, where $(I_j)_{j=1}^{\infty}$ is a family of pairwise disjoint open intervals in $\mathbb{R}$ such that $|I_j|_1 = \ell_j$ for all $j \geq 1$.

Given a relative fractal drum (A, Ω) in $\mathbb{R}^N$ and for a fixed real number r, we define the *relative upper* and *relative lower r-dimensional Minkowski contents* of (A, Ω), respectively, by

$$\mathcal{M}^{*r}(A, \Omega) := \limsup_{t \to 0^+} \frac{|A_t \cap \Omega|}{t^{N-r}}, \quad \mathcal{M}_*^r(A) := \liminf_{t \to 0^+} \frac{|A_t \cap \Omega|}{t^{N-r}}.$$

The *relative upper* and *relative lower box* (or *Minkowski*) *dimensions* of (A, Ω) are then defined, respectively, by

$$\begin{aligned} \overline{\dim}_B(A, \Omega) &:= \inf\{r \in \mathbb{R} : \mathcal{M}^{*r}(A, \Omega) = 0\}, \\ \underline{\dim}_B(A, \Omega) &:= \inf\{r \in \mathbb{R} : \mathcal{M}_*^r(A, \Omega) = 0\}. \end{aligned} \tag{6.1}$$

It is easy to check that $-\infty \leq \underline{\dim}_B(A, \Omega) \leq \overline{\dim}_B(A, \Omega) \leq N$, and it is shown in [41, 44] that the relative box dimensions can indeed attain arbitrary negative values as well, including $-\infty$ (an obvious example is when $A_\delta \cap \Omega = \emptyset$ for some $\delta > 0$). Intuitively, negative relative box dimensions correspond to the *property of flatness* of the RFD under consideration. If $\dim_B(A, \Omega) = -\infty$, then the RFD (A, Ω) is said to be *infinitely flat*. A nontrivial example of an infinitely flat RFD (A, Ω) in $\mathbb{R}^2$ is given by $A := \{(0,0)\}$ and $\Omega := \{(x, y) \in (0,1)^2 : 0 < y < \mathrm{e}^{-1/x}\}$. Other examples of flat RFDs can be found in [41, 44].

If (A,Ω) is such that $\underline{\dim}_B(A,\Omega) = \overline{\dim}_B(A,\Omega)$, then this common value is denoted by $\dim_B(A,\Omega)$ and is called the *box* (or *Minkowski*) *dimension of* (A,Ω). Moreover, if (A,Ω) is such that, for some $D \in (-\infty, N]$, we have $0 < \mathcal{M}_*^D(A,\Omega) \leq \mathcal{M}^{*D}(A,\Omega) < \infty$ (in particular, then $\dim_B(A,\Omega)$ exists and $D = \dim_B(A,\Omega)$), we say that (A,Ω) *is Minkowski nondegenerate*. If $\mathcal{M}_*^D(A,\Omega) = \mathcal{M}^{*D}(A,\Omega)$, then the common value is denoted by $\mathcal{M}^D(A,\Omega)$ and called the *Minkowski content of* (A,Ω). Finally, assuming that (A,Ω) is such that $\mathcal{M}^D(A,\Omega)$ exists and $0 < \mathcal{M}^D(A,\Omega) < \infty$, we say that the RFD (A,Ω) *is Minkowski measurable*.

To any given RFD (A,Ω) in $\mathbb{R}^N$, we can associate the corresponding *relative distance zeta function* and the *relative tube zeta function* defined, respectively, by

$$\zeta_A(s,\Omega) := \int_\Omega d(x,A)^{s-N}\mathrm{d}x, \quad \tilde{\zeta}_A(s,\Omega) := \int_0^\delta t^{s-N-1}|A_t \cap \Omega|\,\mathrm{d}t, \tag{6.2}$$

for all $s \in \mathbb{C}$ with $\mathrm{Re}\, s$ sufficiently large, where δ is a fixed positive real number. They are a valuable theoretical and technical new tool in the study of fractals.

The basic result dealing with relative distance zeta functions, analogous to Theorem 2.2 of §2, is provided by the following theorem.

Theorem 6.1 ([41, 44]) *Let* (A,Ω) *be an arbitrary RFD. Then*:

(*a*) *The distance zeta function* $\zeta_A(\,\cdot\,,\Omega)$ *is holomorphic on* $\{\mathrm{Re}\, s > \overline{\dim}_B(A,\Omega)\}$. *Moreover,* $\Pi(\zeta_A(\,\cdot\,,\Omega)) = \{\mathrm{Re}\, s > \overline{\dim}_B(A,\Omega)\}$*; that is,*

$$D(\zeta_A(\,\cdot\,,\Omega)) = \overline{\dim}_B(A,\Omega). \tag{6.3}$$

(*b*) *If the box* (*or Minkowski*) *dimension* $D := \dim_B(A,\Omega)$ *exists,* $D < N$*, and* $\mathcal{M}_*^D(A,\Omega) > 0$*, then* $\zeta_A(s,\Omega) \to +\infty$ *as* $s \in \mathbb{R}$ *converges to* D *from the right. In particular,* $\mathcal{H}(\zeta_A(\,\cdot\,,\Omega)) = \Pi(\zeta_A(\,\cdot\,,\Omega)) = \{\mathrm{Re}\, s > \dim_B(A,\Omega)\}$*; that is,*

$$D_{\mathrm{hol}}(\zeta_A(\,\cdot\,,\Omega)) = D(\zeta_A(\,\cdot\,,\Omega)) = \dim_B(A,\Omega). \tag{6.4}$$

An entirely analogous result holds for the tube zeta function $\tilde{\zeta}_A(\,\cdot\,,\Omega)$, except for the fact that the hypothesis $D < N$ is no longer needed in the counterpart of part (*b*) of Theorem 6.1.

A very useful property of relative distance zeta functions is the following *scaling property*: for any RFD (A,Ω) and for any positive real number λ, we have

$$\zeta_{\lambda A}(s,\lambda\Omega) = \lambda^s \zeta_A(s,\Omega). \tag{6.5}$$

We refer the interested reader to [41, Chap. 4] and [44–47] for many other related results, examples and comments. We mention, in particular, that 'fractal drums' (that is, 'drums with fractal boundary', in the sense of [22–24], for example)[22] correspond

[22]See also [52, §12.5], [27] and [47] for many other references on fractal drums.

to RFDs of the form $(\partial\Omega, \Omega)$, where Ω is a nonempty bounded open subset of $\mathbb{R}^N$, and that the results discussed in Sect. 5 above are applied in a crucial way in order to show the optimality of certain inequalities pertaining to the meromorphic continuations of the spectral zeta functions of fractal drums (viewed as RFDs); see [41, §4.3] and [47].

7 Fractal Tube Formulas and a Minkowski Measurability Criterion

In this section, we briefly explain how under suitable growth conditions on the relative distance (or tube) zeta function (see a variant of the languidity (resp., of the strong languidity) condition of [52, §5.3] given in [45, 46]), it is possible to recover a pointwise or distributional fractal tube formula for a relative fractal drum (A, Ω) in $\mathbb{R}^N$, expressed as a sum of residues over its visible complex dimensions. These fractal tube formulas, along with a Tauberian theorem due to Wiener and Pitt (which generalizes Ikehara's Tauberian theorem, see [19, 55]) make it possible to derive a Minkowski measurability criterion for a large class of relative fractal drums (and compact subsets) of $\mathbb{R}^N$. These results generalize to higher dimensions the corresponding ones obtained for fractal strings (that is, when $N = 1$) in [52, §8.1 and §8.3].

The results of this section are announced in [45] and fully proved in [46]. (See also [41, Chap. 5].) Furthermore, we refer the interested reader to [41, 46] and [52, §8.2 and §13.1] for additional references on tube formulas in various settings, including [4, 8, 11, 18, 29–31, 34–37, 50, 51, 57, 65, 66]. (See also [52, §13.1, §13.2 and §13.4].)

In order to be able to state the fractal tube formulas, we introduce the following notions, adapted from [52] to the present much more general context. The *screen* S is the graph of a bounded, real-valued, Lipschitz continuous function $S(\tau)$, with the horizontal and vertical axes interchanged: $S := \{S(\tau) + \mathbb{i}\tau \ : \ \tau \in \mathbb{R}\}$ and we let $\sup S := \sup_{\tau\in\mathbb{R}} S(\tau) \in \mathbb{R}$. Given a relative fractal drum (A, Ω) of $\mathbb{R}^N$, we always assume that the screen S lies to the left of the critical line $\{\mathrm{Re}\, s = \overline{\dim}_B(A, \Omega)\}$, i.e., that $\sup S \leq \overline{\dim}_B(A, \Omega)$. Furthermore, the *window* W is defined as $W := \{s \in \mathbb{C} : \mathrm{Re}\, s \geq S(\mathrm{Im}\, s)\}$. The relative fractal drum (A, Ω) is said to be *admissible* if its tube (or distance) zeta function can be meromorphically extended to an open connected neighborhood of some window W.

Assume now that (A, Ω) is an admissible relative fractal drum of $\mathbb{R}^N$ for some screen S such that its distance zeta function satisfies appropriate growth conditions (see [45, 46] for details).[23] Then its relative tube function satisfies the following

[23]Roughly speaking, $\zeta_{(A,\Omega)} := \zeta_A(\,\cdot\,, \Omega)$ is assumed to grow at most polynomially along the vertical direction of the screen and along suitable horizontal directions (avoiding the poles of $\zeta_{(A,\Omega)}$); see [52, Def. 5.2] for the so-called "languidity condition".

identity, for all positive real numbers t sufficiently small[24]:

$$|A_t \cap \Omega| = \sum_{\omega \in \mathcal{P}(\zeta_A(\,\cdot\,,\Omega),W)} \operatorname{res}\left(\frac{t^{N-s}}{N-s}\zeta_A(s,\Omega),\omega\right) + \mathcal{R}(t). \tag{7.1}$$

The above *fractal tube formula* is interpreted pointwise or distributionally, depending on the growth properties of $\zeta_A(\,\cdot\,,\Omega)$ and then, $\mathcal{R}(t)$ is a pointwise or distributional[25] asymptotic error term of order at most $O(t^{N-\sup S})$ as $t \to 0^+$. Moreover, if S lies strictly to the left of the vertical line $\{\operatorname{Re} s = \sup S\}$ (that is, if $S(\tau) < \sup S$, for all $\tau \in \mathbb{R}$), then $\mathcal{R}(t)$ is $o(t^{N-\sup S})$, pointwise or distributionally, as $t \to 0^+$. In the case when $\zeta_A(\,\cdot\,,\Omega)$ satisfies stronger growth assumptions (i.e., the analog of the "strong languidity condition" of [52, Def. 5.3]), we obtain a tube formula without an error term (i.e., $\mathcal{R}(t) \equiv 0$) and with $W = \mathbb{C}$. Following [52], the resulting formula is then called an *exact* fractal tube formula.

The tube formula (7.1) can also be expressed in terms of the relative tube zeta function when analogous growth conditions are imposed on $\tilde{\zeta}_A(\,\cdot\,,\Omega)$[26]:

$$|A_t \cap \Omega| = \sum_{\omega \in \mathcal{P}(\tilde{\zeta}_A(\,\cdot\,,\Omega),W)} \operatorname{res}\left(t^{N-s}\tilde{\zeta}_A(s,\Omega),\omega\right) + \mathcal{R}(t). \tag{7.2}$$

In fact, the key observation for deriving the above formula is the fact that

$$\tilde{\zeta}_A(s,\Omega) = \int_0^{+\infty} t^{s-N-1}\chi_{(0,\delta)}(t)|A_t \cap \Omega|\,\mathrm{d}t = \{\mathfrak{M}f\}(s), \tag{7.3}$$

where χ_E is the characteristic function of the set E, $\{\mathfrak{M}\psi\}(s) := \int_0^{+\infty} t^{s-1}\psi(t)\,\mathrm{d}t$ is the Mellin transform of the function ψ, and $f(t) := t^{-N}\chi_{(0,\delta)}(t)|A_t \cap \Omega|$. One then applies the inverse Mellin transform (see [62]) to recover the relative tube function $t \mapsto |A_t \cap \Omega|$ and proceeds in a similar manner as in [52, Chap. 5] for the case of fractal strings.

As an application, the following result generalizes the Minkowski measurability criterion given in [52, Thm. 8.15] for fractal strings to the present case of relative fractal drums.

Theorem 7.1 (Minkowski measurability criterion, [45, 46]) *Let (A,Ω) be an admissible relative fractal drum of $\mathbb{R}^N$ such that $D := \dim_B A$ exists and $D < N$. Furthermore, assume that its relative distance (or tube) zeta function satisfies*

[24]The ranges within which the formulas are valid are fully specified in [45, 46].

[25]For the precise definition of distributional asymptotics, see [52, §5.4.2], [45, 46] and the relevant references therein.

[26]Note that in light of the functional equation (2.13), assuming growth conditions for ζ_A is essentially equivalent to assuming them for $\tilde{\zeta}_A$ (and vice versa).

appropriate growth conditions[27] *for a screen passing between the critical line* $\{\mathrm{Re}\, s = D\}$ *and all the complex dimensions of A with real part strictly less than D. Then, the following statements are equivalent*:

(*a*) *A is Minkowski measurable.*
(*b*) *D is the only pole of the distance zeta function* ζ_A *located on the critical line* $\{\mathrm{Re}\, s = D\}$*, and it is simple.*

There exist relative fractal drums which do not satisfy the hypothesis of Theorem 7.1 concerning the screen; see [52, Exple. 5.32]. We point out that the fractal tube formula (7.1) can be used to recover (or obtain for the first time) the (relative) fractal tube formulas for a variety of well-known (and not necessarily self-similar) fractal sets, as is illustrated by the following examples.

Example 3 Recall from Proposition 3.3 that the distance zeta function of the Sierpiński carpet A is given for all $s \in \mathbb{C}$ by

$$\zeta_A(s) = \frac{8}{2^s s(s-1)(3^s-8)} + 2\pi\frac{\delta^s}{s} + 4\frac{\delta^{s-1}}{s-1},$$

for $\delta > 1/6$, and is meromorphic on all of $\mathbb{C}$. It is easy to check that ζ_A satisfies growth conditions which are good enough for (7.1) to hold pointwise without an error term and for all $t \in (0, 1/2)$:

$$|A_t| = \sum_{\omega \in \mathcal{P}(\zeta_A, \mathbb{C})} \mathrm{res}\left(\frac{t^{2-s}}{2-s}\zeta_A(s), \omega\right). \tag{7.4}$$

Now, also recall from Proposition 3.3 that $\mathcal{P}(\zeta_A, \mathbb{C}) = \{0, 1\} \cup \{s_k : k \in \mathbb{Z}\}$, where $s_k = \log_3 8 + \frac{2\pi}{\log 3}k\mathbb{i}$ for all $k \in \mathbb{Z}$. Furthermore, $\mathrm{res}(\zeta_A, 0) = 2\pi + 8/7$, $\mathrm{res}(\zeta_A, 1) = 16/5$ and the residues at s_k are given in (3.16); so that (7.4) becomes the following exact, pointwise fractal tube formula, valid for all $t \in (0, 1/2)$:

$$|A_t| = \frac{t^{2-\log_3 8}}{\log 3}\sum_{k=-\infty}^{+\infty}\frac{2^{-s_k} t^{-\frac{2\pi}{\log 3}k\mathbb{i}}}{s_k(s_k-1)(2-s_k)} + \frac{16}{5}t + \left(2\pi + \frac{8}{7}\right)t^2. \tag{7.5}$$

The above example can be generalized to an N-dimensional analog of the Sierpiński carpet (see [41, 46]). We next establish the special case of this assertion for the relative 3-dimensional Sierpiński carpet.

Example 4 Let A be the three-dimensional analog of the Sierpiński carpet and Ω the closed unit cube in $\mathbb{R}^3$. More precisely, we construct A by dividing Ω into 27 congruent cubes and remove the open middle cube, then we iterate this step with each of the 26 remaining smaller closed cubes; and so on, ad infinitum. By choosing

[27] See [45, 46] for details about these growth conditions.

$\delta > 1/6$, we deduce that ζ_A is meromorphic on $\mathbb{C}$ and given for all $s \in \mathbb{C}$ by (see [41, 46])

$$\zeta_A(s, \Omega) = \frac{48 \cdot 2^{-s}}{s(s-1)(s-2)(3^s - 26)}. \tag{7.6}$$

In particular, $\mathcal{P}(\zeta_A(\cdot\,, \Omega), \mathbb{C}) = \{0, 1, 2\} \cup \big(\log_3 26 + \mathbf{p}\mathbb{i}\mathbb{Z}\big)$, where $\mathbf{p} := 2\pi/\log 3$. Furthermore, we have that

$$\operatorname{res}(\zeta_A(\cdot\,,\Omega),0) = -\frac{24}{25}, \quad \operatorname{res}(\zeta_A(\cdot\,\Omega),1) = \frac{24}{23}, \quad \operatorname{res}(\zeta_A(\cdot\,,\Omega),2) = -\frac{6}{17}$$

and, by letting $\omega_k := \log_3 26 + \mathbf{p}k\mathbb{i}$ for all $k \in \mathbb{Z}$,

$$\operatorname{res}(\zeta_A(\cdot\,,\Omega),\omega_k) = \frac{24 \cdot 2^{-\omega_k}}{13 \cdot \omega_k(\omega_k - 1)(\omega_k - 2)\log 3}.$$

Again, the relative distance zeta function $\zeta_A(\cdot\,, \Omega)$ satisfies sufficiently good growth conditions, which enables us to obtain the following exact pointwise relative tube formula, valid for all $t \in (0, 1/2)$:

$$|A_t \cap \Omega| = \frac{24\, t^{3-\log_3 26}}{13 \log 3} \sum_{k=-\infty}^{+\infty} \frac{2^{-\omega_k} t^{-\mathbf{p}k\mathbb{i}}}{(3-\omega_k)(\omega_k - 1)(\omega_k - 2)\omega_k} - \frac{6}{17}t + \frac{12}{23}t^2 - \frac{8}{25}t^3.$$

In particular, we conclude that $\dim_B(A, \Omega) = \log_3 26$ and, by Theorem 7.1, that, as expected, (A, Ω) is not Minkowski measurable.

One can similarly recover the well-known fractal tube formula for the Sierpiński gasket obtained in [35] (and also, more recently, by a somewhat different method, in [4]), as well as a tube formula for its N-dimensional analog described in [41, Chap. 5].[28] We also point out that, in light of the functional equation (2.4), the above fractal tube formulas (7.1) and (7.2) generalize the corresponding ones obtained for fractal strings (i.e., when $N = 1$) in [52, §8.1]. Furthermore, these tube formulas can also be applied to a variety of fractal sets that are not self-similar, including 'fractal nests' and 'geometric chirps' (see [41, Chaps. 3 and 4] for the definitions of these notions and [46] along with [41, Chap. 5] for the actual fractal tube formulas).

We conclude this section by briefly explaining how these results can also be applied in order to recover (and extend) the tube formulas for *self-similar sprays* generated by a suitable bounded open set $G \subset \mathbb{R}^N$. (See [35, 36].) A self-similar spray is a collection $(G_k)_{k\in\mathbb{N}}$ of pairwise disjoint sets $G_k \subset \mathbb{R}^N$, with $G_0 := G$ and such that G_k is a scaled copy of G by some factor $\lambda_k > 0$. The sequence $(\lambda_k)_{k\in\mathbb{N}}$ is

[28] We can also recover and extend the significantly more general fractal tube formulas obtained (for fractal sprays and self-similar tilings) in [36] and used, in particular, in [37].

called the *scaling sequence* associated with the spray and is obtained from a "ratio list" $\{r_1, r_2, \ldots, r_J\}$, with $0 < r_j < 1$ for each $j \in \{1, 2, \ldots, J\}$, by building all possible words based on the ratios r_j. Let now (A, Ω) be the relative fractal drum such that $A := \partial(\cup_{k\in\mathbb{N}} G_k)$ and $\Omega := \cup_{k\in\mathbb{N}} G_k$, with $\overline{\dim}_B(\partial G, G) < N$. Then, it is clear that its relative distance zeta function $\zeta_A(\,\cdot\,, \Omega)$ satisfies the following functional equation, for all $s \in \mathbb{C}$ with $\operatorname{Re} s$ sufficiently large:

$$\zeta_A(s, \Omega) = \zeta_{\partial G}(s, G) + \zeta_{r_1 A}(s, r_1\Omega) + \cdots + \zeta_{r_J A}(s, r_J\Omega), \tag{7.7}$$

where $(r_j A, r_j\Omega)$ denotes the relative fractal drum (A, Ω) scaled by the factor r_j. Furthermore, by using the scaling property (6.5) of the relative distance zeta function, the above equation becomes

$$\zeta_A(s, \Omega) = \zeta_{\partial G}(s, G) + r_1^s \zeta_A(s, \Omega) + \cdots + r_J^s \zeta_A(s, \Omega), \tag{7.8}$$

which yields that

$$\zeta_A(s, \Omega) = \frac{\zeta_{\partial G}(s, G)}{1 - \sum_{j=1}^J r_j^s}. \tag{7.9}$$

It is now enough to assume that the relative distance zeta function $\zeta_{\partial G}(s, G)$ of the generating relative fractal drum $(\partial G, G)$ satisfies suitable growth conditions in order to obtain the following formula for the 'inner' volume of $A_t = (\partial\Omega)_t$ relative to $\Omega := \cup_{k\in\mathbb{N}} G_k$, for all positive t sufficiently small:

$$|A_t \cap \Omega| = \sum_{\omega \in \mathfrak{D}(W) \cup \mathcal{P}(\zeta_{\partial G}(\,\cdot\,, G), W)} \operatorname{res}\left(\frac{t^{N-s} \zeta_{\partial G}(s, G)}{(N-s)\Big(1 - \sum_{j=1}^J r_j^s\Big)}, \omega \right) + \mathcal{R}(t), \tag{7.10}$$

where $\mathfrak{D}(W)$ denotes the set of all visible complex solutions of $\sum_{j=1}^J r_j^s = 1$ (in W) and W is the window defined earlier. It is easy to check that (at least) in the case of monophase or pluriphase generators (in the sense of [35] and [36, 37]), these growth conditions are satisfied, so that one obtains exactly the same distributional or pointwise fractal tube formulas as in [35] or [36], respectively, after having calculated the distance zeta function $\zeta_{\partial G}(\,\cdot\,, G)$ of the generator. Moreover, if $\zeta_{\partial G}(\,\cdot\,, G)$ is strongly languid, we can let $\mathcal{R}(t) \equiv 0$ and $W = \mathbb{C}$ in (7.10) and therefore obtain *exact* fractal tube formulas.

We conclude this survey by pointing out that a broad variety of open problems and suggestions for directions for future research in this area are proposed in [41, Chap. 6].

References

1. A. Baker, *Transcendental Number Theory* (Cambridge University Press, Cambridge, 1975)
2. A.S. Besicovitch, S.J. Taylor, On the complementary intervals of a linear closed set of zero Lebesgue measure. J. Lond. Math. Soc. **29**, 449–459 (1954)
3. G. Bouligand, Ensembles impropres et nombre dimensionnel. Bull. Sci. Math. (2) **52**, 320–344 and 361–376 (1928)
4. A. Deniz, Ş. Koçak, Y. Özdemir, A.E. Üreyen, Tube volumes via functional equations. J. Geom. **105**, 1–10 (2014)
5. K.E. Ellis, M.L. Lapidus, M.C. Mackenzie, J.A. Rock, Partition zeta functions, multifractal spectra, and tapestries of complex dimensions, in *Benoît Mandelbrot*: *A Life in Many Dimensions*, ed. by M. Frame, N. Cohen. The Mandelbrot Memorial Volume (World Scientific, Singapore, 2015), pp. 267–322. (Also: e-print, `arXiv:1007.1467v2 [math-ph]`, 2011; IHES preprint, IHES/M/12/15, 2012.)
6. K.J. Falconer, On the Minkowski measurability of fractals. Proc. Am. Math. Soc. **123**, 1115–1124 (1995)
7. K.J. Falconer, *Fractal Geometry*: *Mathematical Foundations and Applications*, 3rd edn. (John Wiley, Chichester, 2014)
8. H. Federer, Curvature measures. Trans. Am. Math. Soc. **93**, 418–491 (1959)
9. H. Federer, *Geometric Measure Theory* (Springer, New York, 1969)
10. A.O. Gel'fond, *Transcendental and Algebraic Numbers*, Dover Phoenix editions (Dover Publications, New York, 1960)
11. A. Gray, *Tubes*. Progress in Mathematics, vol. 221, 2nd edn. (Birkhäuser, Boston, 2004)
12. B.M. Hambly, M.L. Lapidus, Random fractal strings: their zeta functions, complex dimensions and spectral asymptotics. Trans. Am. Math. Soc. **358**(1), 285–314 (2006)
13. R. Harvey, J. Polking, Removable singularities of solutions of linear partial differential equations. Acta Math. **125**, 39–56 (1970)
14. C.Q. He, M.L. Lapidus, Generalized Minkowski Content, Spectrum of Fractal Drums, Fractal Strings and the Riemann Zeta-Function. Mem. Am. Math. Soc. **127**(608), 1–97 (1997)
15. H. Herichi, M.L. Lapidus, *Quantized Number Theory, Fractal Strings and the Riemann Hypothesis*: *From Spectral Operators to Phase Transitions and Universality*. Research Monograph (World Scientific Publishers, Singapore, 2016, to appear), approx. 220 p.
16. H. Herichi, M.L. Lapidus, Fractal complex dimensions, Riemann hypothesis and invertibility of the spectral operator, in *Fractal Geometry and Dynamical Systems in Pure and Applied Mathematics I: Fractals in Pure Mathematics*, ed. by D. Carfi, M.L. Lapidus, E.P.J. Pearse, M. van Frankenhuijsen. Contemporary Mathematics, vol. 600 (American Mathematical Society, Providence), pp. 51–89. (Also: e-print, `arXiv:1210.0882v3 [math.FA]`, 2013; IHES preprint, IHES/M/12/25, 2012.)
17. L. Horvat, D. Žubrinić, Maximally singular Sobolev functions. J. Math. Anal. Appl. **304**(2), 531–541 (2005)
18. D. Hug, G. Last, W. Weil, A local Steiner-type formula for general closed sets and applications. Math. Z. **246**, 237–272 (2004)
19. J. Korevaar, *Tauberian Theory*: *A Century of Developments* (Springer, Heidelberg, 2004)
20. N. Lal, M.L. Lapidus, Hyperfunctions and spectral zeta functions of Laplacians on self-similar fractals. J. Phys. A: Math. Theor. **45**, (2012), 365205, 14 pp. (Also: e-print, `arXiv:1202.4126v2 [math-ph]`, 2012; IHES preprint, IHES/M/12/14, 2012.)
21. N. Lal, M.L. Lapidus, The decimation method for Laplacians on fractals: spectra and complex dynamics, in *Fractal Geometry and Dynamical Systems in Pure and Applied Mathematics II*: *Fractals in Applied Mathematics*, ed. by D. Carfi, M.L. Lapidus, E.P.J. Pearse, M. van Flankenhuijsen. Contemporary Mathematics, vol. 601 (American Mathematical Society, Providence, 2013), pp. 227–249. (Also: e-print, `arXiv:1302.4007v2 [math-ph]`, 2014; IHES preprint, IHES/M/12/31, 2012.)

22. M.L. Lapidus, Fractal drum, inverse spectral problems for elliptic operators and a partial resolution of the Weyl–Berry conjecture. Trans. Am. Math. Soc. **325**, 465–529 (1991)
23. M.L. Lapidus, Spectral and fractal geometry: from the Weyl–Berry conjecture for the vibrations of fractal drums to the Riemann zeta-function, in *Differential Equations and Mathematical Physics*, ed. by C. Bennewitz; Proceedings of the Fourth UAB International Conference, Birmingham, March 1990 (Academic, New York, 1992), pp. 151–182
24. M.L. Lapidus, Vibrations of fractal drums, the Riemann hypothesis, waves in fractal media, and the Weyl–Berry conjecture, in *Ordinary and Partial Differential Equations*, ed. by B.D. Sleeman, R.J. Jarvis, vol. IV; Proceedings of Twelfth International Conference, Dundee, June 1992. Pitman Research Notes in Mathematics Series, vol. 289 (Longman Scientific and Technical, London, 1993), pp. 126–209
25. M.L. Lapidus, *In Search of the Riemann Zeros*: *Strings, Fractal Membranes and Noncommutative Spacetimes* (American Mathematical Society, Providence, 2008)
26. M.L. Lapidus, Towards quantized number theory: spectral operators and an asymmetric criterion for the Riemann hypothesis (No. 2047, **373** (2015), 24 pp. doi: 10.1098/rsta.2014.0240). (To appear in the special issue of the *Philosophical Transactions of the Royal Society*, Ser. A, titled "*Geometric Concepts in the Foundations of Physics*", 2015.) (Also: e-print, `arXiv:1501.05362v1 [math-ph]`, 2015.)
27. M.L. Lapidus, The sound of fractal strings and the Riemann hypothesis (2015, in press). (To appear in *Analytic Number Theory*: *In Honor of Helmut Maier's 60th Birthday*, ed. by C.M. Pomerance, T. Rassias (Springer, Berlin/Heidelberg/New York, 2015)). (Also: e-print, `arXiv:1505.01548v1 [math-ph]`, 2015.)
28. M.L. Lapidus, J. Lévy-Véhel, J.A. Rock, Fractal strings and multifractal zeta functions. Lett. Math. Phys. **88**(1), 101–129 (2009). (special issue dedicated to the memory of Moshe Flato). (Springer Open Access: DOI 10.1007/s1105-009-0302-y.) (Also: e-print, `arXiv:math-ph/0610015v3`, 2009.)
29. M.L. Lapidus, H. Lu, The geometry of p-adic fractal strings: a comparative survey, in *Advances in Non-Archimedean Analysis*; Proceedings of 11th International Conference on *"p-Adic Functional Analysis"*, Clermont-Ferrand, July 2010, ed. by J. Araujo, B. Diarra, A. Escassut. Contemporary Mathematics, vol. 551 (American Mathematical Society, Providence, 2011), pp. 163–206. (Also: e-print, `arXiv:1105.2966v1 [math.MG]`, 2011.)
30. M.L. Lapidus, H. Lu, M. van Frankenhuijsen, Minkowski measurability and exact fractal tube formulas for p-adic self-similar strings, in *Fractal Geometry and Dynamical Systems in Pure and Applied Mathematics I*: *Fractals in Pure Mathematics*, ed. by D. Carfi, M.L. Lapidus, E.P.J. Pearse, M. van Frankenhuijsen. Contemporary Mathematics, vol. 600 (American Mathematical Society, Providence, 2013), pp. 161–184. (Also: e-print, `arXiv:1209.6440v1 [math.MG], 2012; IHES/M/12/23, 2013.)`
31. M.L. Lapidus, H. Lu, M. van Frankenhuijsen, Minkowski dimension and explicit tube formulas for p-adic fractal strings (2015, preprint)
32. M.L. Lapidus, H. Maier, Hypothèse de Riemann, cordes fractales vibrantes et conjecture de Weyl–Berry modifiée. C. R. Acad. Sci. Paris Sér. I Math. **313**, 19–24 (1991)
33. M.L. Lapidus, H. Maier, The Riemann hypothesis and inverse spectral problems for fractal strings. J. Lond. Math. Soc. (2) **52**, 15–34 (1995)
34. M.L. Lapidus, E.P.J. Pearse, A tube formula for the Koch snowflake curve, with applications to complex dimensions. J. Lond. Math. Soc. **74**(2), 397–414 (2006). (Also: e-print, `arXiv:math-ph/0412029v2`, 2005.)
35. M.L. Lapidus, E.P.J. Pearse, Tube formulas and complex dimensions of self-similar tilings. Acta Applicandae Mathematicae **112**(1), 91–137 (2010). (Springer Open Access: DOI 10.1007/S10440-010-9562-x.) (Also: e-print, `arXiv:math.DS/0605527v5`, 2010; IHES preprint, IHES/M/08/27, 2008.)
36. M.L. Lapidus, E.P.J. Pearse, S. Winter, Pointwise tube formulas for fractal sprays and self-similar tilings with arbitrary generators. Adv. Math. **227**, 1349–1398 (2011). (Also: e-print, `arXiv:1006.3807v3` [math.MG], 2011.)

37. M.L. Lapidus, E.P.J. Pearse, S. Winter, Minkowski measurability results for self-similar tilings and fractals with monophase generators, in *Fractal Geometry and Dynamical Systems in Pure and Applied Mathematics I*: *Fractals in Pure Mathematics*, ed. by D. Carfi, M.L. Lapidus, E.P.J. Pearse, M. van Frankenhuijsen. Contemporary Mathematics, vol. 600 (American Mathematical Society, Providence, 2013), pp. 185–203. (Also: e-print, `arXiv:1104.1641v3 [math.MG]`, 2012; IHES preprint, IHES/M/12/33, 2012.)
38. M.L. Lapidus, C. Pomerance, Fonction zêta de Riemann et conjecture de Weyl–Berry pour les tambours fractals. C. R. Acad. Sci. Paris Sér. I Math. **310**, 343–348 (1990)
39. M.L. Lapidus, C. Pomerance, The Riemann zeta-function and the one-dimensional Weyl–Berry conjecture for fractal drums. Proc. Lond. Math. Soc. (3) **66**(1), 41–69 (1993)
40. M.L. Lapidus, C. Pomerance, Counterexamples to the modified Weyl–Berry conjecture on fractal drums. Math. Proc. Camb. Philos. Soc. **119**, 167–178 (1996)
41. M.L. Lapidus, G. Radunović, D. Žubrinić, *Fractal Zeta Functions and Fractal Drums*: *Higher-Dimensional Theory of Complex Dimensions*. Research Monograph (Springer, New York, 2016, to appear), approx. 530pp.
42. M.L. Lapidus, G. Radunović, D. Žubrinić, Distance and tube zeta functions of fractals and arbitrary compact (2015, preprint)
43. M.L. Lapidus, G. Radunović, D. Žubrinić, Complex dimensions of fractals and meromorphic extensions of fractal zeta functions (2015, in preparation)
44. M.L. Lapidus, G. Radunović, D. Žubrinić, Zeta functions and complex dimensions of relative fractal drums: theory, examples and applications (2015, in preparation)
45. M.L. Lapidus, G. Radunović, D. Žubrinić, Fractal tube formulas and a Minkowski measurability criterion for a class of compact subsets of Euclidean spaces (2014, preprint). (Also: e-print, `arXiv:1411.5733.v2 [math-ph]`, 2014.)
46. M.L. Lapidus, G. Radunović, D. Žubrinić, Fractal tube formulas for compact sets and relative fractal drums, application to a Minkowski measurability criterion (2015, preprint)
47. M.L. Lapidus, G. Radunović, D. Žubrinić, Fractal zeta functions and complex dimensions of relative fractal drums. J. Fixed Point Theory Appl. **15**(2), 321–378 (2014). Festschrift issue in honor of Haim Brezis' 70th birthday. (doi:10.1007/s11784-014-0207-y.) (Also: e-print, `arXiv:1407.8094v3 [math-ph]`, 2014.)
48. M.L. Lapidus, J.A. Rock, Towards zeta functions and complex dimensions of multifractals, Complex Var. Elliptic Equ. **54**(6), 545–560 (2009). (Also: e-print, `arXiv:math-ph/0810.0789`, 2008.)
49. M.L. Lapidus, J.A. Rock, D. Žubrinić, Box-counting fractal strings, zeta functions, and equivalent forms of Minkowski dimension, in *Fractal Geometry and Dynamical Systems in Pure and Applied Mathematics I*: *Fractals in Pure Mathematics*, ed. by D. Carfi, M.L. Lapidus, E.P.J. Pearse, M. van Frankenhuijsen. Contemporary Mathematics, vol. 600 (American Mathematical Society, Providence, 2013), pp. 239–271. (Also: e-print, `arXiv:1207.6681v2 [math-ph]`, 2013; IHES preprint, IHES/M/12/22, 2012.)
50. M.L. Lapidus, M. van Frankenhuijsen, *Fractal Geometry and Number Theory*: *Complex Dimensions of Fractal Strings and Zeros of Zeta Functions* (Birkhäuser, Boston, 2000)
51. M.L. Lapidus, M. van Frankenhuijsen, *Fractal Geometry, Complex Dimensions and Zeta Functions*: *Geometry and Spectra of Fractal Strings*. Springer Monographs in Mathematics (Springer, New York, 2006)
52. M.L. Lapidus, M. van Frankenhuijsen, *Fractal Geometry, Complex Dimensions and Zeta Functions*: *Geometry and Spectra of Fractal Strings*, second revised and enlarged edition (of the 2006 edn., [51]), Springer Monographs in Mathematics (Springer, New York, 2013)
53. J. Lévy-Véhel, F. Mendivil, Multifractal and higher-dimensional zeta functions. Nonlinearity **24**(1), 259–276 (2011)
54. P. Mattila, *Geometry of Sets and Measures in Euclidean Spaces*: *Fractals and Rectifiability* (Cambridge University Press, Cambridge, 1995)
55. A.G. Postnikov, *Tauberian Theory and its Applications*, Proc. Steklov Institute of Mathematics, vol. 1979 (English transl., issue 2, 1980) (American Mathematical Society, Providence, 1980)

56. J. Rataj, S. Winter, Characterization of Minkowski measurability in terms of surface area. J. Math. Anal. Appl. **400**, 120–132 (2013). (Also: e-print, `arXiv:1111.1825v2 [math.CA]`, 2012.)
57. R. Schneider, *Convex Bodies*: *The Brunn–Minkowski Theory* (Cambridge University Press, Cambridge, 2003)
58. J.-P. Serre, *A Course in Arithmetic*, English transl., (Springer, Berlin, 1973)
59. L.L. Stachó, On the volume function of parallel sets. Acta Sci. Math. **38**, 365–374 (1976)
60. A. Teplyaev, Spectral zeta functions of symmetric fractals, in *Progress in Probability*, vol. 57 (Birkhäuser-Verlag, Basel, 2004), pp. 245–262
61. A. Teplyaev, Spectral zeta functions of fractals and the complex dynamics of polynomials. Trans. Am. Math. Soc. **359**, 4339–4358 (2007). (Also: e-print, `arXiv:math.SP/0505546`, 2005.)
62. E.C. Titchmarsh, *Introduction to the Theory of Fourier Integrals*, 2nd edn. (Oxford University Press, Oxford, 1948)
63. C. Tricot, *Curves and Fractal Dimension* (Springer, Berlin, 1995)
64. I.M. Vinogradov (ed.), *Matematicheskaya entsiklopediya*, vol. 2 (Sov. Entsiklopediya, Moscow, 1979), pp. 818–819. (Also: online, public wiki monitored by an editorial board under the management of the European Mathematical Society, URL: www.encyclopediaofmath.org)
65. H. Weyl, On the volume of tubes. Am. J. Math. **61**, 461–472 (1939)
66. M. Zähle, Curvatures and currents for unions of sets with positive reach. Geom. Dedicata **23**, 155–171 (1987)
67. D. Žubrinić, Analysis of Minkowski contents of fractal sets and applications. Real Anal. Exch. **31**(2), 315–354 (2005/2006)
68. D. Žubrinić, Hausdorff dimension of singular sets of Sobolev functions and applications, in *More Progress in Analysis*; Proceedings of the 5th International ISAAC Congress, ed. by H.G.W. Begher, F. Nicolosi (World Scientific, Singapore, 2009), pp. 793–802
69. D. Žubrinić, V. Županović, Fractal analysis of spiral trajectories of some planar vector fields. Bulletin des Sciences Mathématiques **129**(6), 457–485 (2005)

Part IV
Multifractal Theory

Inverse Problems in Multifractal Analysis

Julien Barral

Abstract We present recent results regarding the construction of positive measures with a prescribed multifractal nature, as well as their counterpart in multifractal analysis of Hölder continuous functions.

Keywords Multifractal formalism • Multifractal analysis • Hausdorff dimension • Packing dimension • Large deviations • Inverse problems

Mathematics Subject Classification (2010). Primary 28A78, Secondary 60F10

1 Foreword

This contribution deals with questions different from those considered by the author in his talk, which concerned joint work in progress with Stéphane Seuret on the multifractal nature of Choquet capacities obtained from Gibbs measures via percolation. The results presented here concern the construction of measures and functions with prescribed multifractal nature. Results for measures, the related comments and the sketch of proof given in Sect. 2 are extracted from [3]. Their application to multifractal analysis of functions constitutes the original part of the present paper, developed in Sect. 3.

J. Barral (✉)
Laboratoire d'Analyse, Géométrie et Applications (UMR 7539), Institut Galilée, Université Paris 13, Sorbonne-Paris-Cité, 99 avenue Jean-Baptiste Clément, 93430 Villetaneuse, France
e-mail: barral@math.univ-paris13.fr

C. Bandt et al. (eds.), *Fractal Geometry and Stochastics V*, Progress in Probability 70, DOI 10.1007/978-3-319-18660-3_14

2 Inverse Problems in Multifractal Analysis of Measures

2.1 Generalities About Multifractal Analysis

Let $\mathcal{M}_c^+(\mathbb{R}^d)$ stand for the set of compactly supported Borel positive and finite measures on $\mathbb{R}^d$ ($d \geq 1$). The upper box dimension of a bounded set $E \subset \mathbb{R}^d$ will be denoted $\overline{\dim}_B E$, and its Hausdorff and packing dimensions will be denoted by $\dim_H E$ and $\dim_P E$ respectively (see [20, 44, 52] for definitions).

Multifractal analysis is designed to finely describe geometrically the heterogeneity in the distribution at small scales of the elements of $\mathcal{M}_c^+(\mathbb{R}^d)$. If $\mu \in \mathcal{M}_c^+(\mathbb{R}^d)$, this heterogeneity can be described via the lower and upper local dimensions of μ, namely

$$\underline{d}(\mu, x) = \liminf_{r \to 0^+} \frac{\log(\mu(B(x,r)))}{\log(r)} \quad \text{and} \quad \overline{d}(\mu, x) = \limsup_{r \to 0^+} \frac{\log(\mu(B(x,r)))}{\log(r)},$$

and the level sets

$$\underline{E}(\mu,\alpha)=\Big\{x \in \text{supp}(\mu) : \underline{d}(\mu,x) = \alpha\Big\}, \quad \overline{E}(\mu,\alpha)=\Big\{x \in \text{supp}(\mu) : \overline{d}(\mu,x) = \alpha\Big\},$$

and

$$E(\mu,\alpha) = \underline{E}(\mu,\alpha) \cap \overline{E}(\mu,\alpha) \quad (\alpha \in \mathbb{R} \cup \{\infty\}),$$

where $B(x,r)$ and $\text{supp}(\mu)$ stand for the closed ball of radius $r > 0$ centered at x and the topological support of μ respectively.

The *lower Hausdorff spectrum* of μ is the mapping defined as

$$\underline{f}_\mu^H : \alpha \in \mathbb{R} \cup \{\infty\} \mapsto \dim_H \underline{E}(\mu,\alpha),$$

with the convention that $\dim_H \emptyset = -\infty$, so that $\underline{f}_\mu^H(\alpha) = -\infty$ if $\alpha < 0$. This spectrum provides a geometric hierarchy between the sets $\underline{E}(\mu,\alpha)$, which partition $\text{supp}(\mu)$. Here, the lower local dimension is emphasized as it provides at any point the best pointwise Hölder control one can have on the measure μ at small scales. However, the upper local dimension is of course of interest, and much attention is paid in general, especially in ergodic theory, to the sets $E(\mu,\alpha)$ of points at which one has an exact local dimension. The *Hausdorff spectrum* of μ is the mapping defined as

$$f_\mu^H : \alpha \in \mathbb{R} \cup \{\infty\} \mapsto \dim_H E(\mu,\alpha).$$

After basic observations made by physicists [26, 27], mathematicians derived, and in many cases justified, the heuristic according to which, for a measure possessing a self-conformal like property, f_μ^H should be the Legendre transform of a kind of free energy function, called the L^q-spectrum. This led to an abundant literature on what has become called multifractal formalism (see e.g. [12, 13, 39, 48, 49, 52]).

To be more specific we need some definitions. Given $I \in \{\mathbb{R}, \mathbb{R}\cup\{\infty\}\}$ and $f : I \to \mathbb{R}\cup\{-\infty\}$, the domain of f is defined as $\mathrm{dom}(f) = \{x \in I : f(x) > -\infty\}$. For $\tau : \mathbb{R} \to \mathbb{R}\cup\{-\infty\}$, if $\mathrm{dom}(\tau) \neq \emptyset$, the concave Legendre-Fenchel transform of τ is the upper-semi continuous concave function defined as $\tau^* : \alpha \in \mathbb{R} \mapsto \inf\{\alpha q - \tau(q) : q \in \mathrm{dom}(\tau)\}$ (see [54]). If, moreover, $0 \in \mathrm{dom}(\tau)$, we define its (extended) concave Legendre-Fenchel transform as

$$\tau^* : \alpha \in \mathbb{R}\cup\{\infty\} \mapsto \begin{cases} \inf\{\alpha q - \tau(q) : q \in \mathrm{dom}(\tau)\} & \text{if } \alpha \in \mathbb{R}, \\ \inf\{\alpha q - \tau(q) : q \in \mathrm{dom}(\tau) \cap \mathbb{R}_-\} & \text{if } \alpha = \infty, \end{cases}$$

with the conventions $\infty \times q = -\infty$ if $q < 0$ and $\infty \times 0 = 0$. Consequently, $\infty \in \mathrm{dom}(\tau^*)$ if and only if $0 = \min(\mathrm{dom}(\tau))$, and in this case $\tau^*(\infty) = -\tau(0) = \max(\tau^*)$. In any case, τ^* is upper semi-continuous over $\mathrm{dom}(\tau^*)$, and concave over the interval $\mathrm{dom}(\tau^*) \setminus \{\infty\}$ (here the notion of upper semi-continuous function is relative to $\mathbb{R}\cup\{\infty\}$ endowed with the topology generated by the open subsets of $\mathbb{R}$ and the sets $(\alpha, \infty) \cup \{\infty\}$, $\alpha \in \mathbb{R}$).

Now, define the *L^q-spectrum* of $\mu \in \mathcal{M}_c^+(\mathbb{R}^d)$ as

$$\tau_\mu : q \in \mathbb{R} \mapsto \liminf_{r\to 0^+} \frac{\log \sup \left\{ \sum_i \mu(B(x_i, r))^q \right\}}{\log(r)},$$

where the supremum is taken over all the centered packings of $\mathrm{supp}(\mu)$ by closed balls of radius r. The following properties are standard and proved for instance in [39].

Proposition 2.1 *Let $\mu \in \mathcal{M}_c^+(\mathbb{R}^d)$.*

1. *τ_μ is concave and non-decreasing; $\tau_\mu(1) = 0$, $-d \le \tau_\mu(0) = -\overline{\dim}_B \,\mathrm{supp}(\mu)$ ≤ 0.*
2. *Either $\mathrm{dom}(\tau_\mu) = \mathbb{R}$, or $\mathrm{dom}(\tau_\mu) = \mathbb{R}_+$, according to whether the exponent $\limsup_{r\to 0^+} \dfrac{\log(\inf\{\mu(B(x,r)) : x \in \mathrm{supp}(\mu)\})}{\log(r)}$ is finite or not. Moreover τ_μ^* is non-negative on its domain, which is a closed subinterval of $\mathbb{R}_+ \cup \{\infty\}$.*

For $\alpha \in \mathbb{R}$ we always have (see [39, Section 3] or [49, Section 2.7])

$$f_\mu^H(\alpha) \le \underline{f}_\mu^H(\alpha) \le \tau_\mu^*(\alpha) \le \max(\alpha, -\tau_\mu(0)) \le \max(\alpha, d); \tag{2.1}$$

we also have

$$f_\mu^H(\infty) \le \tau_\mu^*(\infty)$$

(see [3]), a dimension equal to $-\infty$ meaning that the set is empty. Notice that due to (2.1), if $f_\mu^H(\alpha) \geq \alpha$ at some α, then $0 \leq \alpha \leq d$ and $f_\mu^H(\alpha) = \tau_\mu^*(\alpha) = \alpha$, so that α is a fixed point of τ_μ^*. Moreover, since $\tau_\mu(1) = 0$ and τ_μ is concave, the set of fixed points of τ_μ^* is the interval $[\tau_\mu'(1^+), \tau_\mu'(1^-)]$.

We say that μ obeys the multifractal formalism at $\alpha \in \mathbb{R} \cup \{\infty\}$ if $\underline{f}_\mu^H(\alpha) = \tau_\mu^*(\alpha)$, and that the multifractal formalism holds (globally) for μ if it holds at all $\alpha \in \mathbb{R} \cup \{\infty\}$. If $\underline{f}_\mu^H(\alpha)$ can be replaced by $f_\mu^H(\alpha)$ in the previous definition, we say that the multifractal formalism holds strongly. In this case one has

$$\dim_H E(\mu,\alpha) = \dim_P E(\mu,\alpha) = \dim_H \underline{E}(\mu,\alpha) = \dim_H \overline{E}(\mu,\alpha) = \tau_\mu^*(\alpha).$$

The multifractal formalism turns out to hold globally, or on some non-trivial subinterval of $\mathrm{dom}(\tau_\mu^*)$, for some important classes of continuous measures, namely some classes of self-conformal measures (including certain Bernoulli convolutions), Gibbs and weak Gibbs measures on hyperbolic dynamical systems (see e.g. [13, 16, 21–25, 39, 42, 52, 53] and [3] for more references), and scale invariant limits of certain multiplicative chaos [1, 6, 17, 29, 46]; in these cases it also holds strongly. It also holds for some natural classes of discrete measures (see e.g. [2, 9, 34, 51] as well as references in [3]). Other examples are special self-affine or Gibbs measures on self-affine Sierpinski carpets [4, 7, 38, 50], or on almost all the attractors of IFS associated with certain families of $d \times d$ invertible matrices with small enough singular values [5, 18, 19], as well as generic probability measures on a compact subset of $\mathbb{R}^d$ [11, 14].

The measures mentioned above share the geometric property of being exact dimensional, i.e. for such a measure μ, there exists $D \in [0, d]$ such that $\underline{d}(\mu, x) = \overline{d}(\mu, x) = D$, μ-almost everywhere. This implies $D \in [\tau_\mu'(1^+), \tau_\mu'(1^-)]$ and μ strongly obeys the multifractal formalism at D. In fact, for any $\mu \in \mathcal{M}_c^+(\mathbb{R}^d)$, for μ-almost every x one has $\tau_\mu'(1^+) \leq \underline{d}(\mu, x) \leq \overline{d}(\mu, x) \leq \tau_\mu'(1^-)$ ([48]), and for most of the continuous measures mentioned above, $\tau_\mu'(1)$ exists, hence equals D; also, τ_μ is piecewise C^1, and even analytic in certain cases, a typical example being Gibbs measures associated with Hölder potentials on repellers of $C^{1+\alpha}$ conformal mappings. Another property of these measures is that, when they obey the multifractal formalism globally, they are homogeneously multifractal (HM), in the sense that the lower Hausdorff spectrum of the restriction of μ to any closed ball whose interior intersects $\mathrm{supp}(\mu)$ is equal to the lower Hausdorff spectrum of μ.

2.2 Full Illustration of the Multifractal Formalism

Theorem 2.2 ([3]) *Let $\tau : \mathbb{R} \to \mathbb{R} \cup \{-\infty\}$ be a concave function satisfying the necessary properties (see Proposition 2.1) to be the L^q-spectrum of some element of $\mathcal{M}_c^+(\mathbb{R}^d)$. Let $D \in [\tau'(1^+), \tau'(1^-)]$. There exists an (HM) measure $\mu \in \mathcal{M}_c^+(\mathbb{R}^d)$,* *exact dimensional with dimension D, and which strongly satisfies the multifractal formalism with* $\tau_\mu = \tau$.

Remark 1 In [3] we develop much more general results by using a finer multifractal formalism to prescribe and distinguish Hausdorff and packing dimensions of the level sets $\{x \in \text{supp}(\mu) : \underline{d}(\mu, x) = \alpha, \ \overline{d}(\mu, x) = \beta\}$, $(\alpha \leq \beta \leq \infty)$. The connection with Olsen's multifractal formalism [49] is also studied.

It is interesting to complete this statement by describing the possible behaviors of (τ_μ, τ_μ^*) (see Figs. 1–6). For this we need to extend the notion of Legendre-Fentchel transform to functions $f : \mathbb{R}\cup\{\infty\} \to \mathbb{R}\cup\{-\infty\}$: for such an f, if $\text{dom}(f)\cap\mathbb{R} \neq \emptyset$, we define the concave Legendre-Fenchel transform of f as

$$f^* : q \in \mathbb{R} \mapsto \inf\{q\alpha - f(\alpha) : \alpha \in \text{dom}(f)\},$$

with the conventions $q \times \infty = \text{sign}(q) \times \infty$ if $q \neq 0$ and $0 \times \infty = 0$. Consequently, if $\infty \in \text{dom}(f)$ and f is bounded from above, then $0 = \min(\text{dom}(f^*))$ and $f^*(0) = -\max(\sup(f)_{|\mathbb{R}}), f(\infty))$; moreover, f^* is concave over $\text{dom}(f^*)$, upper semi-continuous over $\text{dom}(f^*) \setminus \{0\}$, and upper semi-continuous at 0 only if and only if $f(\infty) = \max(f)$.

Proposition 2.3 ([3, 39]) *Suppose that* $\tau : \mathbb{R} \to \mathbb{R} \cup \{-\infty\}$ *satisfies the properties of the* L^q*-spectrum described in Proposition* 2.1. *One has* $(\tau^*)^* = \tau$ *on* $\mathbb{R}$*, and:*

1. *If* $\text{dom}(\tau) = \mathbb{R}$*, then* $\text{dom}(\tau^*)$ *is the compact interval* $I = [\tau'(\infty), \tau'(-\infty)]$*,* τ^* *is concave and continuous on its domain.*

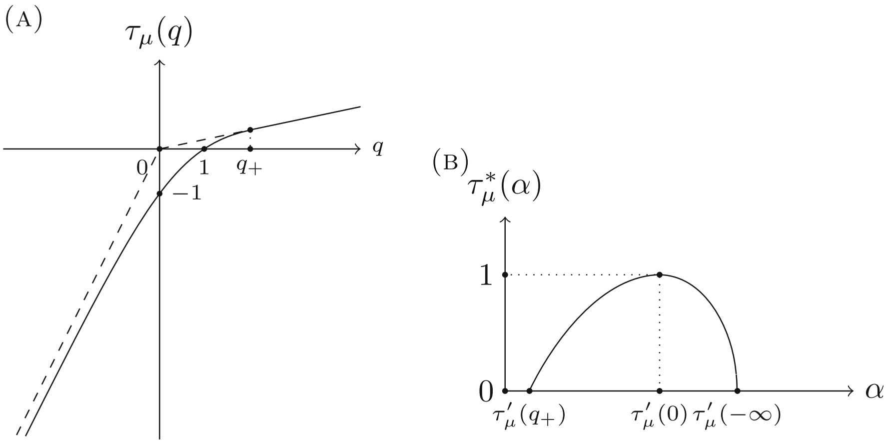


Fig. 1 Illustration of Proposition 2.3.1. when the domain of τ_μ is a non trivial interval and τ_μ is differentiable, with a second order phase transition at some $q_+ > 1$.
The case of a trivial interval $\{\alpha_0\}$ would correspond to a monofractal measure with $0 \leq \alpha_0 \leq d$, $\tau_\mu(q) = \alpha_0(q-1)$ for all $q \in \mathbb{R}$, $\tau_\mu^*(\alpha) = \alpha$ if $\alpha = \alpha_0$ and $\tau_\mu^*(\alpha) = -\infty$ otherwise. (**a**) The L^q spectrum of μ. (**b**) Its Legendre transform

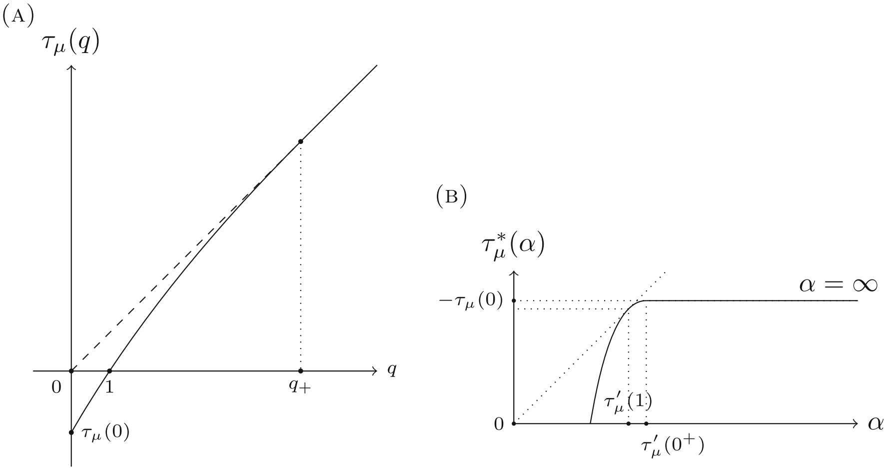


Fig. 2 Illustration of Proposition 2.3.2(b) when $\tau'_\mu(0^+) < \infty$, τ_μ is differentiable, and it has a second order phase transition at some $q_+ > 1$. (**a**) The L^q spectrum of μ. (**b**) Its Legendre transform

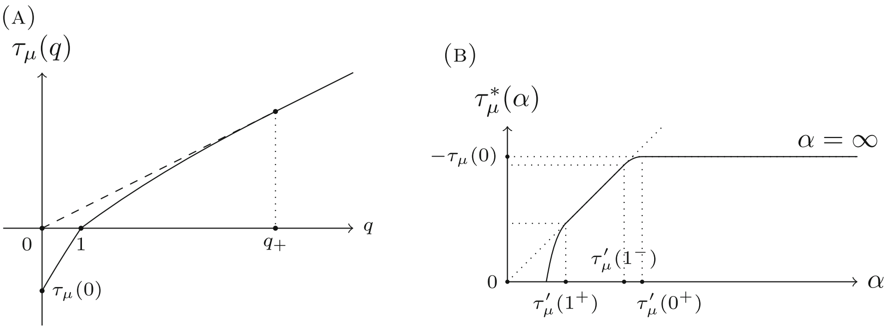


Fig. 3 Illustration of Proposition 2.3.2(b) when $\tau'_\mu(0^+) < \infty$, τ_μ is not differentiable at 1, and it has a second order phase transition at some $q_+ > 1$. (**a**) The L^q spectrum of μ. (**b**) Its Legendre transform

2. *If dom*$(\tau) = \mathbb{R}_+$, *then* $\infty \in$ *dom*(τ^*) *with* $\tau^*(\infty) = -\tau(0)$ *and:*

   (a) *If* $\tau(0) = 0$ *then* $\tau = 0$ *over* $\mathbb{R}_+$, *dom*$(\tau^*) = \mathbb{R}_+ \cup \{\infty\}$ *and* $\tau^* = 0$ *over* $\mathbb{R}_+ \cup \{\infty\}$.

   (b) *If* $\tau(0) < 0$ *and* τ *is continuous at* 0^+, *then dom*$(\tau^*) = [\tau'(\infty), \infty]$, τ^* *is concave, continuous, and increasing over* $[\tau'(\infty), \tau'(0^+))$, $\tau^*(\alpha) = -\tau(0) = \tau^*(\infty) = -\tau(0)$ *for all* $\alpha \in [\tau'(0^+), \infty)$ *and* τ^* *is continuous at* ∞; *there are two distinct behaviors according to whether* $\tau'(0^+) < \infty$ *or not.*

   (c) *If* $\tau(0) < 0$ *and* τ *is discontinuous at* 0^+, *then dom*$(\tau^*) = [\tau'(\infty), \infty]$. *Moreover, for all* $\alpha \in [\lim_{q\to 0^+} \tau'(q^-), \infty)$ *one has* $\tau^*(\alpha) = -\tau(0^+) < \tau^*(\infty) = -\tau(0)$, *so that* τ^* *is concave and continuous on* $[\tau'(\infty), \infty)$ *and*

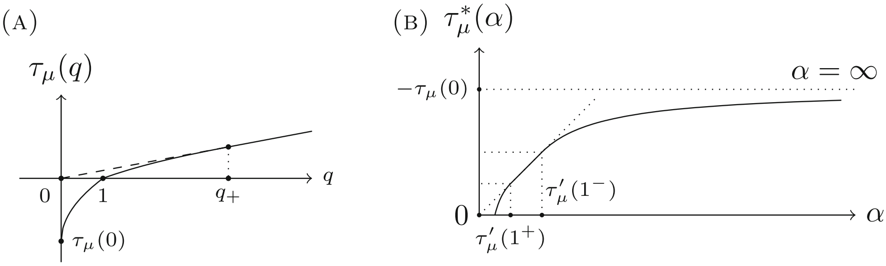


Fig. 4 Illustration of Proposition 2.3.2(b) when $\tau'_\mu(0^+) = \infty$, τ_μ is not differentiable at 1, and it has a second order phase transition at some $q_+ > 1$. (**a**) The L^q spectrum of μ. (**b**) Its Legendre transform

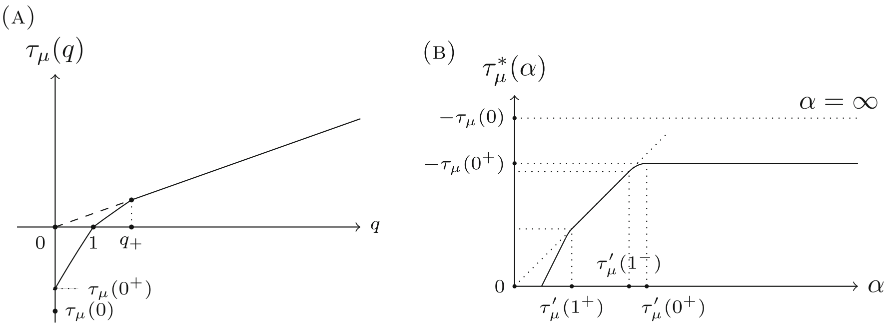


Fig. 5 Illustration of Proposition 2.3.2(c) when $\tau'_\mu(0^+) < \infty$, τ_μ is not differentiable at 1, and it has another first order phase transition at some $q_+ > 1$. (**a**) The L^q spectrum of μ. (**b**) Its Legendre transform

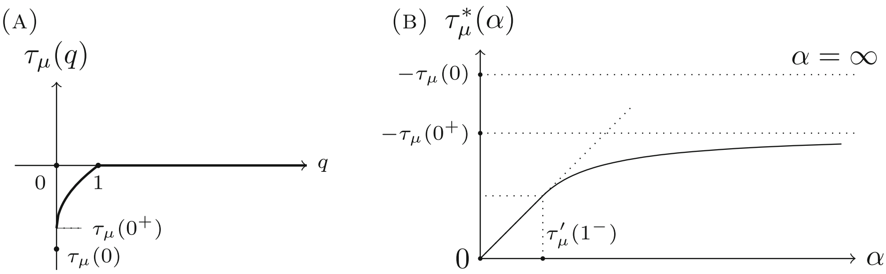


Fig. 6 Illustration of Proposition 2.3.2(c) when $\tau'_\mu(0^+) = \infty$, τ_μ is not differentiable at 1 and $\tau'_\mu(1^+)$ takes the minimal value 0. (**a**) The L^q spectrum of μ. (**b**) Its Legendre transform

discontinuous at ∞ *(there are also two cases, according to* $\lim_{q\to 0^+} \tau'(q^-)$ *equals* ∞ *or not).*

Remark 2 The behavior described in Proposition 2.3.1 is illustrated, for instance, by (weak) Gibbs measures on conformal repellers [25, 49, 52]. The behaviors described by Proposition 2.3.2(b) are illustrated by some Gibbs measures on countable

Markov shifts and their geometric realizations [30, 42, 43], which also obey the multifractal formalism, though in [30, 43] the set $E(\mu, \infty)$ is not studied. The fact that the behaviors described in Proposition 2.3.2(a) and (c) be illustrated by measures obeying the mutifractal formalism seems to be new.

Remark 3 In [28], when $d = 1$, for each $D \in (0, 1)$ one finds an exact dimensional measure μ with dimension D and L^q-spectrum equal to $\min(q-1, 0)$ over $\mathbb{R}_+$. It is also worth mentioning that in [10] one finds examples of inhomogeneous Bernoulli measures over $[0, 1]$ with an L^q-spectrum presenting countably many points of non-differentiability over $[1, +\infty)$.

2.3 Measures with prescribed lower Hausdorff spectrum

In general, $\mathrm{dom}(f^H_{-\mu}) = \{\alpha \in \mathbb{R} \cup \{\infty\} : \underline{E}(\mu, \alpha) \neq \emptyset\}$ is not a closed subinterval of $[0, \infty]$, and even when it is the case, the restriction of $f^H_{-\mu}$ to $\mathrm{dom}(f^H_{-\mu}) \cap \mathbb{R}_+$ is not necessarily concave. Consequently, it is also natural to study the inverse problem consisting of associating to a function $f : \mathbb{R} \cup \{\infty\} \to [0, d] \cup \{-\infty\}$ whose domain is a subset of $\mathbb{R}_+ \cup \{\infty\}$ and such that $f(\alpha) \leq \alpha$ for all $\alpha \geq 0$, an (HM) measure whose lower Hausdorff spectrum is equal to f. In [3] we construct such a measure μ, exact dimensional, when f shares important properties with τ^*_μ; specifically, f is taken in the family:

$$\mathcal{F}(d) = \left\{ f : \mathbb{R} \cup \{\infty\} \to [0, d] \cup \{-\infty\} : \begin{cases} \mathrm{dom}(f) \text{ is a closed subset of } [0, \infty] \\ f \text{ is u.s.c., } \mathrm{Fix}(f) \neq \emptyset \\ f(\alpha) \leq \alpha \text{ for all } \alpha \in \mathrm{dom}(f) \end{cases} \right\},$$

where $\mathrm{Fix}(f)$ $(\subset [0, d])$ stands for the set of fixed points of f.

Theorem 2.4 ([3]) *Let $f \in \mathcal{F}(d)$. For each $D \in \mathrm{Fix}(f)$, there exists an (HM) measure $\mu \in \mathcal{M}^+_c(\mathbb{R}^d)$, exact dimensional with dimension D, such that $f^H_{-\mu} = f$.*

Remark 4

(1) The measures constructed in the proofs of Theorems 2.2 and 2.4 are continuous and supported on Cantor sets.
(2) Our approach does not make it possible to replace $f^H_{-\mu} = f$ by $f^H_\mu = f$ in the previous statement unless $\mathrm{dom}(f) = \mathrm{Fix}(f)$ or $\mathrm{dom}(f)$ is an interval and f is concave over $\mathrm{dom}(f) \cap \mathbb{R}_+$. It turns out that the proof given in [3] can be slightly improved so that μ is absolutely continuous with respect to Lebesgue measure when $D = d$.

Remark 5 (Related result by Z. Buczolich and S. Seuret) The prescription of the lower Hausdorff spectrum has also been studied in [15]. The authors work on $\mathbb{R}$ and construct (HM) continuous measures, not exact dimensional, but with upper Hausdorff dimension equal to 1, and whose support is equal to $[0, 1]$. Moreover, the lower Hausdorff spectrum is prescribed in the class $\mathcal{F}$ of functions $f : \mathbb{R}_+ \to [0, 1] \cup \{-\infty\}$ which satisfy: $f(1) = 1$, $\mathrm{dom}(f)$ is a closed subinterval of $[0, 1]$ of the form $[\alpha, 1]$ such that $\alpha > 0$, and $f_{|[\alpha,1)} = \max(g_{|[\alpha,1)}, 0)$, where the function g has the following properties: (i) g is the supremum of a sequence of functions $(g_n)_{n\geq 1}$, such that each g_n is constant over its domain supposed to be a closed subinterval of $[0, 1]$ and $g_n(\beta) \leq \beta$ for all $\beta \in [0, 1]$; (ii) $[\alpha, 1]$ is the smallest closed interval containing the support of g.

It is also shown that for an (HM) measure to be supported by the whole interval $[0, 1]$, it is necessary that the support of its lower Hausdorff spectrum contains an interval of the form $[\alpha, 1]$, $(0 \leq \alpha \leq 1)$.

The authors also study the case of non-(HM) measures. They construct measures that are non exact dimensional with upper Hausdorff dimension 1 whose support is equal to $[0, 1]$, with a prescribed lower Hausdorff spectrum in the broader class $\tilde{\mathcal{F}}$ of functions f which satisfy that $f(1) = 1$, $0 < \inf(\mathrm{dom}(f))$, and $f_{|\mathrm{dom}(f)\setminus\{1\}} = g_{|\mathrm{dom}(f)\setminus\{1\}}$, where g satisfies property (i). This includes all such functions f for which g is lower semi-continuous. Simultaneously, they also construct a non-(HM) measure with lower Hausdorff spectrum given by g.

Remark 6 The spectra previously defined make sense if measures are replaced by non-negative functions of subsets of $\mathbb{R}^d$ to which a notion of support is associated. This is the case for instance for Choquet capacities. In [40], the prescription of $\alpha \mapsto \dim_H E(C, \alpha)$ is studied, where C is a (HM) Choquet capacity on subsets of $[0, 1]$ but not a positive measure, which makes the situation easier to study; spectra are prescribed in a broader class than $\tilde{F}$, but defined in a similar spirit.

In [41], one finds non-(HM) non-negative functions C of subsets of $[0, 1]$, which are not measures, for which the spectrum $\alpha \mapsto \lim_{\epsilon\to 0^+} \dim_H \bigcup_{s>0} \bigcap_{0<r<s} \{x \in \mathrm{supp}(C) : r^{\alpha+\epsilon} \leq C(B(x, r)) \leq r^{\alpha-\epsilon}\}$ is prescribed in the class of upper semi-continuous functions $f : \mathbb{R}_+ \mapsto [0, 1] \cup \{-\infty\}$ with non-empty compact domain.

2.4 Outline of the Proof of Theorem 2.4

Let us sketch the main ideas leading to the construction of the measure μ provided by Theorem 2.4. To establish Theorem 2.2 one must improve this approach in order to control both the finer level sets $E(\mu, \alpha)$ and the upper large deviations spectrum $\overline{f}_\mu^{\mathrm{LD}}$ of μ when $f = \tau_\mu^*$, and the relation $\tau_\mu = \overline{f}_\mu^{\mathrm{LD}*}$.

For simplicity, we assume that $\mathrm{dom}(f)$ is a non-trivial interval $[\alpha_{\min}, \alpha_{\max}] \subset \mathbb{R}_+$, f is continuous over $[\alpha_{\min}, \alpha_{\max}]$, $0 \leq f(\alpha) \leq \min(\alpha, d)$ over $[\alpha_{\min}, \alpha_{\max}]$, and $f(D) = D$ for a unique point D in $[\alpha_{\min}, \alpha_{\max}]$. The homogeneity of the construction of the measure μ automatically implies that the measure is (HM).

At first one shows (independently of f) that for each $\gamma \in [0, d]$ and $\alpha \geq \gamma$, one can find two Borel probability measures $\mu_{\alpha,\gamma}$ and $\nu_{\alpha,\gamma}$ supported on $[0, 1]^d$ such that $\mu_{\gamma,\gamma} = \nu_{\gamma,\gamma}$, $\nu_{\alpha,\gamma}$ is exact dimensional with dimension γ, and $\nu_{\alpha,\gamma}$ is concentrated on $E(\mu_{\alpha,\gamma}, \alpha)$, as well as on the set defined similarly but with $\alpha(\mu, x)$ replaced by $\lim_{n\to\infty} \frac{\log(\mu(I_n(x)))}{-n\log(2)}$, where $I_n(x)$ stands for the closure of dyadic cube semi-open to the right containing x.

Set $A_1 = \{\alpha_1 = D\}$, and for each integer $m \geq 1$, define $A_{m+1} = A_m \cup \{\alpha_{m+1}\}$, where $\alpha_{m+1} \in [\alpha_{\min}, \alpha_{\max}] \setminus A_m$, in such a way that the set $\{\alpha_m : m \geq 1\}$ is dense in $[\alpha_{\min}, \alpha_{\max}]$. By using the previous property with $\gamma = f(\alpha)$, for all $m \geq 1$ one gets an integer n_m such that for all $\alpha \in A_m$, for all $n \geq n_m$, there is a collection $G_{m,n}(\alpha)$ of about $2^{nf(\alpha)}$ dyadic subcubes of $[0, 1]^d$ such that for all $I \in G_{m,n}(\alpha)$ one has $\mu_{\alpha,f(\alpha)}(I) \approx 2^{-n\alpha}$, $\nu_{\alpha,f(\alpha)}(I) \approx 2^{-nf(\alpha)}$, and $\sum_{I\in G_{m,n}(\alpha)} \nu_{\alpha,f(\alpha)}(I) \in [1/2, 1]$.

For every integer $m \geq 2$, one considers m dyadic closed subcubes $L_{\alpha_1}, \dots, L_{\alpha_m}$ of $[0, 1]^d$, of the same generation n'_m, so that the $2^{-n'_m/5}$ neighborhood of each L_{α_i} does not intersect any of the other L_{α_j}.

The measure μ is constructed on a Cantor set $K = \bigcap_{m\geq 1} \bigcup_{I\in \mathbf{G}_m}$, where the $\mathbf{G}_m$ are families of closed dyadic subcubes of $[0, 1]^d$ of generation g_m tending to ∞ as $m \to \infty$, constructed recursively according to a scheme roughly as follows:

One obtains $\mathbf{G}_1$ by considering the measure $\mu_{\alpha_1,f(\alpha_1)} = \mu_{D,D}$, an integer $N_1 \geq n_1$ much bigger than n'_2 and setting $\mathbf{G}_1 = G_{1,N_1}(\alpha_1) = G_{1,N_1}(D)$. This yields the probability measure μ_1 defined on $\mathbf{G}_1$ as

$$\mu_1(I) = \frac{\mu_{D,D}(I)}{\sum_{I'\in \mathbf{G}_1} \mu_{D,D}(I')}.$$

This measure satisfies $\mu_1(I) \approx 2^{-N_1 D}$. Suppose now that the set $\mathbf{G}_m$ has been constructed, as well as a probability measure μ_m on its elements. One takes $N_{m+1} \geq n_{m+1}$ an integer much bigger than $\max(g_m, n'_{m+2})$, and for each $1 \leq i \leq m + 1$, one considers the measure $\mu_{\alpha_i,f(\alpha_i)}$ and the associated set $G_{m+1}(\alpha_i) := G_{m+1,N_{m+1}}(\alpha_i)$. For each $1 \leq i \leq m + 1$ and $I_m \in \mathbf{G}_m$, one defines the set of the elements of $\mathbf{G}_{m+1}$ contained in I_m as $\bigcup_{i=1}^{m+1} \mathbf{G}_{m+1}(I_m, \alpha_i)$, where $\mathbf{G}_{m+1}(I_m, \alpha_i) = \{I_m \cdot L_{\alpha_i} \cdot I : I \in G_{m+1}(\alpha_i)\}$, and the concatenation $J \cdot J'$ of two closed subcubes of $[0, 1]^d$ is obtained as the cube $f_J(J')$, where f_J is the natural contracting similitude mapping $[0, 1]^d$ onto J (this operation is associative). One gets a probability measure μ_{m+1} on $\mathbf{G}_{m+1}$ by setting, for $I \in G_{m+1}(\alpha_i)$:

$$\mu_{m+1}(I_m \cdot L_{\alpha_i} \cdot I) = \mu_m(I_m) \frac{\mu_{\alpha_i,f(\alpha_i)}(I)}{\sum_{\alpha\in A_{m+1}} \sum_{I'\in G_{m+1}(\alpha)} \mu_{\alpha,f(\alpha)}(I')}. \tag{2.2}$$

This makes it possible to define a Borel probability measure carried on K and coinciding with μ_m over $\mathbf{G}_m$ for all $m \geq 1$.

Since $f(\alpha) < \alpha$ except for $\alpha = \alpha_1 = D$, if N_{m+1} is taken big enough, in (2.2) for each $i > 1$ the contribution of the elements of $G_{m+1}(\alpha)$ is roughly $2^{N_{m+1}(f(\alpha)-\alpha)}$ hence is negligible so that the denominator is equivalent to the single contribution of $\sum_{I' \in G_{m+1}(D)} \mu_{D,D}(I') \in [1/2, 1]$. Consequently, for $I_{m+1} \in \mathbf{G}_{m+1}$ of the form $I_m \cdot L_{\alpha_i} \cdot I$, $I \in G_{m+1}(\alpha_i)$, we have the following estimate:

$$\mu(I_{m+1}) \approx \mu_m(I_m)\mu_{\alpha_i, f(\alpha_i)}(I) \approx \mu_m(I_m)2^{-\alpha_i N_{m+1}} \approx 2^{-\alpha_i g_{m+1}} \tag{2.3}$$

because $g_m \ll N_{m+1}$. Also, we have that $\#G_{m+1}(\alpha_i) \approx 2^{f(\alpha_i)N_{m+1}}$, hence

$$\begin{aligned} &\#\{I \in \mathbf{G}_{m+1} : I \in \mathbf{G}_{m+1}(I_m, \alpha_i) \text{ with } I_m \in \mathbf{G}_m\} \\ &\qquad = (\#\mathbf{G}_m)(\#G_{m+1}(\alpha_i)) \approx 2^{f(\alpha_i)g_{m+1}}, \end{aligned}$$

again because $g_m \ll N_{m+1}$. The previous estimate and the continuity of f essentially yield that f is an upper bound for $\underline{f}^H$. Combined with (2.3), it shows that at generation $m + 1$, the mass of μ is essentially carried by the intervals $I_m \cdot L_D \cdot I$, $I \in G_{m+1}(D)$, since we have $1 = \|\mu\| \approx \sum_{i=1}^{m+1} 2^{f(\alpha_i)g_{m+1}} 2^{-\alpha_i g_{m+1}} = \sum_{i=1}^{m+1} 2^{(f(\alpha_i)-\alpha_i)g_{m+1}} \approx 2^{(f(\alpha_1)-\alpha_1)g_{m+1}} = 1$ (recall that $\alpha_1 = f(\alpha_1) = D$). This can be strengthened to show that μ is exact D-dimensional.

Another important fact is the natural existence of a family of auxiliary measures used to find a sharp lower bound for $\underline{f}^H$: with each $\hat{\beta} = (\beta_m)_{m\geq 1} \in \prod_{m=1}^{\infty} A_m$ is associated the Cantor subset of K defined as $K_{\hat{\beta}} = \bigcap_{m\geq 1} \bigcup_{I \in \mathbf{G}_{\hat{\beta},m}} I$, where $\mathbf{G}_{\hat{\beta},m}$ is the subset of $\mathbf{G}_m$ obtained by selecting only the intervals of the construction for which one considers the exponent $\beta_i \in A_i$ at step i for all $1 \leq i \leq m$. Using (2.3) and finer properties of the measures $\mu_{\alpha,\gamma}$ one can show that $K_{\hat{\beta}} \subset \underline{E}(\mu, \beta)$, where $\beta = \liminf_{m\to\infty} \beta_m$. Moreover, the measures $\nu_{\beta_m, f(\beta_m)}$ can be used to construct a nice auxiliary probability measure $\nu_{\hat{\beta}}$ carried by $K_{\hat{\beta}}$. At first one defines recursively a sequence of measures $(\nu_{\hat{\beta},m})_{m\geq 1}$ on the atoms of the sets $\mathbf{G}_{\hat{\beta},m}$, $m \geq 1$, as follows: $\nu_{\hat{\beta},1}$ is the restriction of $\nu_{D,D}$ to $\mathbf{G}_{\hat{\beta},1} (= \mathbf{G}_1)$, and assuming that $\nu_{\hat{\beta},m}$ is constructed on $\mathbf{G}_{\hat{\beta},m}$, if $I_m \in \mathbf{G}_{\hat{\beta},m}$, for $I \in G_{m+1}(\beta_{m+1})$ one sets

$$\nu_{\hat{\beta},m+1}(I_m \cdot L_{\beta_{m+1}} \cdot I) = \nu_{\hat{\beta},m}(I_m) \frac{\nu_{\beta_{m+1}, f(\beta_{m+1})}(I)}{\sum_{I' \in G_{m+1}(\beta_{m+1})} \nu_{\beta_{m+1}, f(\beta_{m+1})}(I')}.$$

This yields a Borel probability measure $\nu_{\hat{\beta}}$ supported on $K_{\hat{\beta}}$ such that $\nu_{\hat{\beta}}(I_m \cdot L_{\beta_{m+1}} \cdot I) = \nu_{\hat{\beta},m+1}(I_m \cdot L_{\beta_{m+1}} \cdot I) \approx \nu_{\hat{\beta},m}(I_m)\nu_{\beta_{m+1}, f(\beta_{m+1})}(I)$, so that $\nu_{\hat{\beta}}(I_m \cdot L_{\beta_{m+1}} \cdot I) \approx \nu_{\beta_{m+1}, f(\beta_{m+1})}(I) \approx 2^{-f(\beta_{m+1})g_{m+1}}$ (again since $g_m \ll N_{m+1}$). This can be strengthened to $\dim_H(\nu_{\hat{\beta}}) = \liminf_{m\to\infty} f(\beta_m)$, which yields $\dim_H K_{\hat{\beta}} \geq \liminf_{m\to\infty} f(\beta_m)$ by the mass distribution principle (see [20]). Finally, if $\beta \in [\alpha_{\min}, \alpha_{\max}]$ and $\lim_{m\to\infty} \beta_m = \beta$, we get $\underline{f}^H(\beta) = \dim_H \underline{E}(\mu, \beta) \geq f(\beta)$ by continuity of f.

3 Application to Multifractal Analysis of Hölder Continuous Functions

Multifractal analysis of functions has developed in parallel to multifractal analysis of measures, mainly under the impulse of Frisch and Parisi's note about multifractality in fully developed turbulence [26], and with its own multifractal formalisms [33, 35–37, 47]. These are based on the link between pointwise Hölder regularity and the wavelet expansions of Hölder continuous functions [31].

Theorems 2.2 and 2.4 can be used to construct Hölder continuous wavelet series with prescribed upper semi-continuous lower Hausdorff spectra, and also to give a full illustration of the multifractal formalism for Hölder continuous functions based on the wavelet leaders [36], according to the bridge made in [8] between this formalism and the multifractal formalism for measures. We will restrict ourselves to the case $d = 1$.

To be more specific, recall first that if $F : \mathbb{R} \to \mathbb{R}$ is a bounded Hölder continuous function, for each $x_0 \in \mathbb{R}$, one defines the pointwise Hölder exponent of f at x_0 as

$$h_F(x_0) = \sup\{h \geq 0 : \text{for some polynomial}\, P, \\ |F(x) - P(x - x_0)| = O(|x - x_0|^h) \text{ as } |x - x_0| \to 0\},$$

where $|x - x_0|$ stands for the Euclidean norm of $x - x_0$. This exponent is the counterpart for functions of the lower local dimension for measures.

One usually calls the mapping

$$h \mapsto \dim_H\{x \in \mathbb{R} : h_F(x) = h\} \quad (h \in \mathbb{R} \cup \{\infty\})$$

the singularity spectrum of F (we keep the terminology lower Hausdorff spectrum for a slightly different spectrum defined below). Notice that if f is γ-Hölder, then $\{x \in \mathbb{R} : h_F(x) = h\} = \emptyset$ if $h < \gamma$.

We are going to restrict the study to $[0, 1]$. We fix a wavelet basis $\{\psi_I\}$ (I describing all the dyadic subintervals of $\mathbb{R}$), so that the mother wavelet is in the Schwartz class (see [45, Ch. 3]) and the ψ_I are normalized to have the same supremum norm.

Denoting $\{\lambda_I\}$ the collection of the wavelet coefficients of F in the basis $\{\psi_I\}$, let $L_I = \sup\{|\lambda_{I'}|\}$, the supremum being taken over all the dyadic intervals included either in I or in the two dyadic intervals of the same generation as I neighboring I. Then, let supp(F) be the closed set of those $x \in [0, 1]$ such that $|L(I_n(x))| > 0$ for all $n \geq 1$, where $I_n(x)$ stands for the closure of the unique semi-open to the right dyadic cube of generation n which contains x. According to [36], this set does not depend on ψ; moreover, for $x \in \text{supp}(F)$, one has

$$h_F(x) = \liminf_{n\to\infty} \frac{\log_2(|L(I_n(x))|)}{-n}.$$

For $h \in \mathbb{R} \cup \{\infty\}$, we set

$$\underline{E}(F, h) = \{x \in \text{supp}(F) : h_F(x) = h\}.$$

The lower Hausdorff spectrum of F is the mapping

$$\underline{f}_F^H : h \mapsto \dim_H \underline{E}(F, h) \quad (h \in \mathbb{R} \cup \{\infty\}).$$

We say that F is homogeneously multifractal (HM) if for all $h \in \mathbb{R} \cup \{\infty\}$, the Hausdorff dimension of $\underline{E}(F, h) \cap B$ does not depend on the ball B whose interior intersects supp(F).

A basic idea [8] to relate multifractal analysis of functions to that of measures is to consider wavelet series of the form

$$F_{\mu,\gamma_1,\gamma_2} = \sum_{I \subset [0,1]} |I|^{\gamma_1} \mu(I)^{\gamma_2} \psi_I,$$

where $|I|$ stands for the diameter of I, $\gamma_1 \geq 0$, $\gamma_2 > 0$, $\mu \in \mathcal{M}_c^+(\mathbb{R})$ with supp$(\mu) \subset [0, 1]$, and

$$\gamma = \gamma_1 + \gamma_2 \liminf_{n\to\infty} \frac{\log_2(\max\{\mu(I) : I \text{ dyadic } \subset [0, 1],\ |I| = 2^{-n}\})}{-n} > 0,$$

so that the function $F_{\mu,\gamma_1,\gamma_2}$ is β-Hölder continuous for all $0 < \beta < \gamma$. Then, the study achieved in [8] yields

$$\underline{E}(F_{\mu,\gamma_1,\gamma_2}, h) = \underline{E}\Big(\mu, \frac{h - \gamma_1}{\gamma_2}\Big) \tag{3.1}$$

for all $h \in \mathbb{R} \cup \{\infty\}$, so that any information about the multifractal structure of measures should transfer to a similar one for this class of wavelet series. In particular, it is clear from (3.1) that $\dim_H E(F_{\mu,\gamma_1,\gamma_2}, h) \leq \frac{h-\gamma_1}{\gamma_2}$.

3.1 Prescription of the Lower Hausdorff Spectrum

Theorem 3.1 *Let* $f : \mathbb{R}_+ \cup \{\infty\} \to [0, 1] \cup \{-\infty\}$ *be upper semi-continuous. Suppose that* dom(f) *is a closed subset* $\mathcal{I}$ *of* $[0, \infty]$ *such that* $0 < \min(\mathcal{I}) < \infty$. *There exists an (HM) Hölder continuous function* F *such that* $\underline{f}_F^H = f$.

Proof For $\lambda > 0$ set $\theta(\lambda) = \sup\{f(h)/\lambda h : h \in \mathcal{I}\}$, with the convention $x/\infty = 0$ for all $x \geq 0$. Since f is upper semi-continuous and bounded over its domain, $\theta(\lambda)$ is reached at some $h < \infty$. Moreover, the mapping θ is continuous, and we have $\theta(1/\min(\mathcal{I})) \leq 1$ by definition of f. Now we distinguish two cases.

If $f \not\equiv 0$ over $\mathcal{I}$, then $\theta(\lambda)$ tends to ∞ as λ tends to 0^+, so the continuity of θ yields $0 < \lambda_0 \leq 1/\min(\mathcal{I})$ such that $\theta(\lambda_0) = 1$, hence $f(h) \leq \lambda_0 h$ for all $h \in \mathcal{I}$, with equality at some h. Let $\tilde{f} = f(\lambda_0^{-1}\cdot)$. By construction we have $\tilde{f} \in \mathcal{F}(1)$. Put $\tilde{f}$ in Theorem 2.4 to get an (HM) measure in $\mathcal{M}_c^+(\mathbb{R})$ supported on $[0, 1]$ whose lower Hausdorff spectrum is given by $\tilde{f}$. Then $F = F_{\mu,0,\lambda_0^{-1}}$ is Hölder continuous and has f as lower Hausdorff spectrum by (3.1).

If $f \equiv 0$ on $\mathcal{I}$, then $\tilde{f} = f(\cdot - \min(\mathcal{I}))$ belongs to $\mathcal{F}(1)$ (with 0 as unique fixed point). Put $\tilde{f}$ in Theorem 2.4 to get an (HM) measure in $\mathcal{M}_c^+(\mathbb{R})$ supported on $[0, 1]$ whose lower Hausdorff spectrum is given by $\tilde{f}$. Then $F = F_{\mu,\min(\mathcal{I}),1}$ is Hölder continuous and has f as lower Hausdorff spectrum.

Remark 7 In [15], the measures described in Remark 5(1) are used to construct (HM) functions of the form $F = F_{\mu,\gamma_1,\gamma_2}$ with $\mathrm{supp}(F) = [0, 1]$. Previously in [32], S. Jaffard constructed non-(HM) wavelet series with prescribed spectrum in the class of functions $f : (0,\infty) \to [0, 1]$ which are representable as the supremum of a countable collection of step functions.

3.2 Full Illustration of the Multifractal Formalism

Our results also yield a full illustration of the multifractal formalism for Hölder continuous functions whose support is a subset of $[0, 1]$. This requires some preliminary definitions and facts.

If $F = \sum_I \lambda_I \psi_I$ is a non-trivial such function, i.e. $\emptyset \neq \mathrm{supp}(F) \subset [0, 1]$, denote by $T(q)$ the L^q-spectrum associated with the wavelet leaders $(L_I)_{I\subset[0,1]}$, i.e. the concave non-decreasing function

$$T_F(q) = \liminf_{n\to\infty} \frac{-1}{n} \log_2 \sum_{I\in G_n^*} L_I^q \quad (q \in \mathbb{R}),$$

where G_n^* stands for the set of dyadic cubes I of generation n for which $L_I > 0$. Due to [36] again, this function does not depend on the choice of $\{\psi_I\}$ if the mother wavelet is in the Schwartz class. Moreover, if F takes the form $F_{\mu,\gamma_1,\gamma_2}$, one has almost immediately

$$\tau_F(q) = \tau_\mu(\gamma_2 q) - \gamma_1 q \quad (q \in \mathbb{R}). \tag{3.2}$$

From now on we discard the trivial case of $\lim_{n\to\infty} \frac{\log_2(\max\{L_I : I\in G_n^*\})}{-n} = \infty$, so that $T_F = -\overline{\dim}_B\, \mathrm{supp}(F)\mathbf{1}_{\{0\}} + (-\infty)\mathbf{1}_{\mathbb{R}_-^*} + (\infty)\mathbf{1}_{\mathbb{R}_+^*}$ and $\mathrm{supp}(F) = \underline{E}(F,\infty)$.

Now we have $\liminf_{n\to\infty} \frac{\log_2(\max\{L_I : I\in G_n^*\})}{-n} < \infty$, so there exists $\beta \geq 0$ such that $T_F(q) \leq \beta q$ for all $q \geq 0$, which ensures that T_F takes values in $\mathbb{R} \cup \{-\infty\}$.

We say that F satisfies the multifractal formalism if $\dim_H \underline{E}(F, h) = T_F^*(h)$ for all $h \in \mathbb{R}_+ \cup \{\infty\}$. This is essentially the multifractal formalism considered in [36]. One simple, but important, observation in [8] is that from (3.1) and (3.2) follows the fact that if $\mu \in \mathcal{M}_c^+(\mathbb{R})$ is supported on $[0, 1]$ and obeys the multifractal formalism for measures, then if $F_{\mu,\gamma_1,\gamma_2}$ is Hölder continuous, it obeys the multifractal formalism just defined above.

Let us now examine some features of the L^q-spectrum when the multifractal formalism holds. We distinguish three important properties denoted (i)–(iii): since $L_I = O(|I|^\alpha)$ for some $\alpha > 0$ by the Hölder continuity assumption, we have

(i) There exists $\alpha > 0$ and $c \in [0, 1]$ (here $c = \overline{\dim}_B \operatorname{supp}(F) = -T_F(0)$) such that $T_F(q) \geq \alpha q - c$ for all $q \geq 0$. Moreover,

(ii) T_F satisfies the same properties as τ in Proposition 2.1.2, in particular T_F^* is non-negative over its domain.

Due to (i), we can define $q_0 = \inf\{q \geq 0 : T_F(q) > 0\}$. If F satisfies the multifractal formalism, we must have the third property:

(iii) Either $q_0 > 0$, or $T_F'(0^+) > 0$ and $T_F(q) = T_F'(0^+)q$ for all $q > 0$.

Let us justify this fact. If $q_0 = 0$, there exists $c' \in \mathbb{R}_+$ such that $T_F(q) = T_F'(0^+)q + c'$ for all $q > 0$, for otherwise by concavity of T_F one has $-\infty < T_F^*(T_F'(q^-)) < 0$ for all $q > 0$ large enough so that $T_F'(q^-) < T_F'(0^+)$, while T_F^* must be non-negative over its domain. Also, since there exists $\beta > 0$ such that $0 < T_F(q) \leq \beta q$ for $q > 0$, we have $c' = 0$. If, moreover, F satisfies the multifractal formalism, we must have $T_F'(0^+) > 0$, otherwise $T_F = -\overline{\dim}_B \operatorname{supp}(F)$ over $\mathbb{R}_+^*$, and no Hölder continuous function F can fulfill the multifractal formalism with T_F as L^q-spectrum; indeed, this would imply $T_F^*(0) = \overline{\dim}_B \operatorname{supp}(F) \geq 0$ hence $\underline{E}(F_\mu, 0) \neq \emptyset$.

Theorem 3.2 *Suppose that a non-decreasing concave function T satisfies the above properties* (i)–(iii) *necessary to be the* L^q*-spectrum of a Hölder continuous function whose support is a non-empty subset of* $[0, 1]$. *Then there exists an (HM) Hölder continuous function F with* $\operatorname{supp}(F) \subset [0, 1]$, *which satisfies the multifractal formalism with* $T_F = T$.

Proof Let $q_0 = \inf\{q \geq 0 : T(q) > 0\}$. If $q_0 > 0$, then $\tau(q) = T(q_0 q)$ satisfies the properties of Proposition 2.1, so that it is the L^q-spectrum of an exact dimensional measure μ of dimension D, for any $D \in [q_0 T'(q_0^+), q_0 T'(q_0^-)] \subset [0, 1]$ by Theorem 2.2. Moreover, the inequality $\tau_\mu(q) \geq \alpha q_0 q - c$ implies that $\tau_\mu^*(\beta) = -\infty$ for all $\beta < \alpha q_0$. Consequently, the function $F = F_{\mu,0,1/q_0}$ is $(\alpha - \epsilon)$-Hölder continuous for all $\epsilon > 0$, and due to (3.1) and (3.2) it fulfills the multifractal formalism for wavelet leaders with $T_F : q \mapsto \tau_\mu(q/q_0) = T(q)$.

If $q_0 = 0$, the function defined as $\tau(q) = T(q) - T'(0^+)q$ satisfies the conditions required by Proposition 2.1. Take the (HM) measure μ associated with this function τ by Theorem 2.2. Then, the function $F_{\mu,T'(0^+),1}$ is $(T'(0^+) - \epsilon)$-Hölder continuous for all $\epsilon > 0$ and due to (3.1) and (3.2) it fulfills the multifractal formalism for wavelet leaders, with $T_F : q \mapsto \tau_\mu(q) + T'(0^+)q = T(q)$.

Remark 8 In [35], S. Jaffard uses a multifractal formalism associated with wavelet coefficients (not leaders). He introduces a class of concave functions such that to each element τ of this class he can associate a Baire space V built from Besov spaces, so that generically an element of V has a non-decreasing Hausdorff spectrum obtained as the Legendre transform of τ computed by taking the infimum over a subdomain of $\mathbb{R}_+$.

References

1. N. Attia, J. Barral, Hausdorff and packing spectra, large deviations, and free energy for branching random walks in $\mathbb{R}^d$. Commun. Math. Phys. **331**, 139–187 (2014)
2. V. Aversa, C. Bandt, The multifractal spectrum of discrete measures. 18th winter school on abstract analysis. Acta Univ. Carolin. Math. Phys. **31**(2), 5–8 (1990)
3. J. Barral, Inverse problems in multifractal analysis of measures. Ann. Sci. Ec. Norm. Sup. (to appear). arXiv:1311.3895v3
4. J. Barral, D.-J. Feng, Weighted thermodynamic formalism and applications. Asian J. Math. **16**, 319–352 (2012)
5. J. Barral, D.-J. Feng, Multifractal formalism for almost all self-affine measures. Commun. Math. Phys. **318**, 473–504 (2013)
6. J. Barral, B. Mandelbrot, Random multiplicative multifractal measures, in *Fractal Geometry and Applications*. Proceedings of Symposia in Pure Mathematics, vol. 72, Part 2 (American Mathematical Society, Providence, 2004), pp. 3–90
7. J. Barral, M. Mensi, Gibbs measures on self-affine Sierpinski carpets and their singularity spectrum. Ergod. Theory Dyn. Syst. **27**, 1419–1443 (2007)
8. J. Barral, S. Seuret, From multifractal measures to multifractal wavelet series. J. Fourier Anal. Appl. **11**, 589–614 (2005)
9. J. Barral, S. Seuret, The singularity spectrum of Lévy processes in multifractal time. Adv. Math. **214**, 437–468 (2007)
10. A. Batakis, B. Testud, Multifractal analysis of inhomogeneous Bernoulli products. J. Stat. Phys. **142**, 1105–1120 (2011)
11. F. Bayard, Multifractal spectra of typical and prevalent measures. Nonlinearity **26**, 353–367 (2013)
12. F. Ben Nasr, Analyse multifractale de mesures, C. R. Acad. Sci. Paris, Série I **319**, 807–810 (1994)
13. G. Brown, G. Michon, J. Peyrière, On the multifractal analysis of measures. J. Stat. Phys. **66**, 775–790 (1992)
14. Z. Buczolich, S. Seuret, Typical measures on $[0, 1]^d$ satisfy a multifractal formalism. Nonlinearity **23**, 2905–2911 (2010)
15. Z. Buczolich, S. Seuret, Measures and functions with prescribed homogeneous spectrum. J. Fractal. Geom. **1**, 295–333 (2014)
16. G.A. Edgar, R.D. Mauldin, Multifractal decompositions of digraph recursive fractals. Proc. Lond. Math. Soc. **65**, 604–628 (1992)
17. K.J. Falconer, The multifractal spectrum of statistically self-similar measures. J. Theor. Probab. **7**, 681–702 (1994)
18. K.J. Falconer, Generalized dimensions of measures on self-affine sets. Nonlinearity **12**, 877–891 (1999).
19. K.J. Falconer, Representation of families of sets by measures, dimension spectra and Diophantine approximation. Math. Proc. Camb. Philos. Soc. **128**, 111–121 (2000)
20. K.J. Falconer, *Fractal Geometry – Mathematical Foundations and Applications*, 3rd ed. (Wiley, Hoboken, 2014)
21. A.H. Fan, Multifractal analysis of infinite products. J. Stat. Phys. **86**, 1313–1336 (1997)

22. D.-J. Feng, The limit Rademacher functions and Bernoulli convolutions associated with Pisot numbers. Adv. Math. **195**, 24–101 (2005)
23. D.-J. Feng, Gibbs properties of self-conformal measures and the multifractal formalism. Ergod. Theory Dyn. Syst. **27**, 787–812 (2007)
24. D.-J. Feng, K.S. Lau, Multifractal formalism for self-similar measures with weak separation condition. J. Math. Pures Appl. **92**, 407–428 (2009)
25. D.-J. Feng, E. Olivier, Multifractal analysis of the weak Gibbs measures and phase transition–Application to some Bernoulli convolutions. Ergod. Theory Dyn. Syst. **23**, 1751–1784 (2003)
26. U. Frisch, G. Parisi, On the singularity structure of fully developed turbulence, in *Turbulence, and Predictability in Geophysical Fluid Dynamics and Climate Dynamics*, ed. by M. Ghil et al. (North Holland, Amsterdam/New York, 1985), pp. 84–88
27. T. C. Halsey, M. H. Jensen, L.P. Kadanoff, I. Procaccia, B.I. Shraiman, Fractal measures and their singularities: the characterisation of strange sets. Phys. Rev. A **33**, 1141–1151 (1986)
28. Y. Heurteaux, Dimension of measures: the probabilistic approach. Publ. Mat. **51**, 243–290 (2007)
29. R. Holley, E.C. Waymire, Multifractal dimensions and scaling exponents for strongly bounded random fractals. Ann. Appl. Probab. **2**, 819–845 (1992)
30. G. Iommi, Multifractal Analysis for countable Markov shifts. Ergod. Theory Dyn. Syst. **25**, 1881–1907 (2005)
31. S. Jaffard, Exposants de Hölder en des points donnés et coefficients d'ondelettes. C. R. Acad. Sci. Paris Série I **308**, 79–81 (1989)
32. S. Jaffard, Construction de fonctions multifractales ayant un spectre de singularités prescrit. C. R. Acad. Sci. Paris, Sér. I **315**, 19–24 (1992)
33. S. Jaffard, Multifractal formalism for functions. I. Results valid for all functions. II Self-similar functions. SIAM J. Math. Anal. **28**, 944–970 & 971–998 (1997)
34. S. Jaffard, The multifractal nature of Lévy processes. Probab. Theory Relat. Fields **114**, 207–227 (1999)
35. S. Jaffard, On the Frisch-Parisi conjecture. J. Math. Pures Appl. **79**, 525–552 (2000)
36. S. Jaffard, Wavelet techniques in multifractal analysis, in *Fractal Geometry and Applications*. Proceedings of Symposia in Pure Mathematics, vol. 72, Part 2 (American Mathematical Society, Providence, 2004), pp. 91–152
37. S. Jaffard, Y. Meyer, On the pointwise regularity of functions in critical Besov spaces. J. Funct. Anal. **175**, 415–434 (2000)
38. J.F. King, The singularity spectrum for general Sierpinski carpets. Adv. Math. **116**, 1–8 (1995)
39. K.S. Lau, S.-M. Ngai, Multifractal measures and a weak separation condition. Adv. Math. **141**, 45–96 (1999)
40. J. Lévy-Véhel, R. Vojak, Multifractal analysis of Choquet capacities. Adv. Appl. Math. **20**, 1–43 (1998)
41. J. Lévy-Véhel, C. Tricot, On various multifractal spectra, in *Fractal Geometry and Stochastics III*. Progress in Probability, vol. 57 (Birkhäuser, Boston, 2004), pp. 23–42
42. B.B. Mandelbrot, C.J.G. Evertsz, Y. Hayakawa, Exactly self-similar left sided multifractal measures. Phys. Rev. A **42**, 4528–4536 (1990)
43. B.B. Mandelbrot, R. Riedi, Multifractal formalism for infinite multinomial measures. Adv. Appl. Math. **16**, 132–150 (1995)
44. P. Mattila, *Geometry of Sets and Measures in Euclidean Spaces* (Cambridge University Press, Cambridge, 1995)
45. Y. Meyer, *Ondelettes et Opérateurs I* (Hermann, Paris, 1990)
46. G.M. Molchan, Scaling exponents and multifractal dimensions for independent random cascades. Commun. Math. Phys. **179**, 681–702 (1996)
47. J.F. Muzy, A. Arneodo, E. Bacry, A multifractal formalism revisited with wavelets. Int. J. Bifur. Chaos Appl. Sci. Eng. **4**, 245–302 (1994)
48. S.-M. Ngai, A dimension result arising from the L^q-spectrum of a measure. Proc. Am. Math. Soc. **125**, 2943–2951 (1997)
49. L. Olsen, A multifractal formalism. Adv. Math. **116**, 82–196 (1995)

50. L. Olsen, Self-affine multifractal Sierpinski sponges in $\mathbb{R}^d$. Pac. J. Math. **183**, 143–199 (1998)
51. L. Olsen, N. Snigireva, Multifractal spectra of in-homogeneous self-similar measures. Indiana Univ. Math. J. **57**, 1789–1844 (2008)
52. Y. Pesin, *Dimension Theory in Dynamical Systems: Contemporary Views and Applications* (University of Chicago Press, Chicago, 1997)
53. D.A. Rand, The singularity spectrum $f(\alpha)$ for cookie-cutters. Ergod. Theory Dyn. Syst. **9**, 527–541 (1989)
54. R.T. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, 1970)

Multifractal Analysis Based on *p*-Exponents and Lacunarity Exponents

Patrice Abry, Stéphane Jaffard, Roberto Leonarduzzi, Clothilde Melot, and Herwig Wendt

Abstract Many examples of signals and images cannot be modeled by locally bounded functions, so that the standard multifractal analysis, based on the Hölder exponent, is not feasible. We present a multifractal analysis based on another quantity, the p-exponent, which can take arbitrarily large negative values. We investigate some mathematical properties of this exponent, and show how it allows us to model the idea of "lacunarity" of a singularity at a point. We finally adapt the wavelet based multifractal analysis in this setting, and we give applications to a simple mathematical model of multifractal processes: Lacunary wavelet series.

Keywords Scale Invariance • Fractal • Multifractal • Hausdorff dimension • Hölder regularity • Wavelet • Lacunarity exponent • p-exponent

1 Introduction

The origin of fractal geometry can be traced back to the quest for non-smooth functions, rising from a key question that motivated a large part of the progresses in analysis during the nineteenth century: Does a continuous function necessarily have points of differentiability? A negative answer to this question was supplied

P. Abry • R. Leonarduzzi
Signal, Systems and Physics, Physics Dept., CNRS UMR 5672, Ecole Normale Supérieure de Lyon, Lyon, France
e-mail: patrice.abry@ens-lyon.fr; roberto.leonarduzzi@ens-lyon.fr

S. Jaffard (✉)
Université Paris Est, Laboratoire d'Analyse et de Mathématiques Appliquées, CNRS UMR 8050, UPEC, Créteil, France
e-mail: jaffard@u-pec.fr

C. Melot
I2M, CNRS UMR 7373, Université Aix-Marseille, Marseille, France
e-mail: melot@cmi.univ-mrs.fr

H. Wendt
IRIT, CNRS UMR 5505, University of Toulouse, Toulouse, France
e-mail: herwig.wendt@irit.fr

C. Bandt et al. (eds.), *Fractal Geometry and Stochastics V*, Progress in Probability 70, DOI 10.1007/978-3-319-18660-3_15

by Weierstrass when he built his famous counterexamples, now referred to as the *Weierstrass functions*

$$\mathcal{W}_{a,b}(x) = \sum_{n=0}^{+\infty} a^n \cos(b^n \pi x) \tag{1.1}$$

where $0 < a < 1$, b was an odd integer and $ab > 1 + 3\pi/2$. The fact that they are continuous and nowhere differentiable was later sharpened by Hardy in a way which requires the notion of *pointwise Hölder regularity*, which is the most commonly used notion of pointwise regularity in the function setting. We assume in the following that the functions or distributions we consider are defined on $\mathbb{R}$. However, most results that we will investigate extend to several variables.

Definition 1.1 Let $f : \mathbb{R} \to \mathbb{R}$ be a locally bounded function, $x_0 \in \mathbb{R}$ and let $\gamma \geq 0$; f belongs to $C^\gamma(x_0)$ if there exist $C > 0$, $R > 0$ and a polynomial P of degree less than γ such that:

$$\text{for a.e. } \quad x \quad \text{such that} \quad |x - x_0| \leq R, \qquad |f(x) - P(x - x_0)| \leq C|x - x_0|^\gamma. \tag{1.2}$$

The Hölder exponent of f at x_0 is

$$h_f(x_0) = \sup\{\gamma : \ f \ \text{ is } \ C^\gamma(x_0)\}. \tag{1.3}$$

The Hölder exponent of $\mathcal{W}_{a,b}$ is a constant function, which is equal to $H = -\log a / \log b$ at every point (see e.g. [14] for a simple, wavelet-based proof); since $H < 1$ we thus recover the fact that $\mathcal{W}_{a,b}$ is nowhere differentiable, but the sharper notion of Hölder exponent allows us to draw a difference between each of the Weierstrass functions, and classify them using a regularity parameter that takes values in $\mathbb{R}^+$. The graphs of Weierstrass functions supply important examples of fractal sets that still motivate research (the determination of their Hausdorff dimensions remains partly open, see [6]). In applications, such fractal characteristics have been used for classification purposes. For instance, an unorthodox use was the discrimination between Jackson Pollock's original paintings and fakes using the box dimension of the graph supplied by the pixel by pixel values of a high resolution photograph of the painting, see [25].

The status of everywhere irregular functions was, for a long time, only the one of academic counter-examples, such as the Weierstrass functions. This situation changed when stochastic processes like Brownian motion (whose Hölder exponent is $H = 1/2$ everywhere) started to play a key role in the modeling of physical phenomena. Nowadays, experimentally acquired signals that are everywhere irregular are prevalent in a multitude of applications, so that the classification and modeling of such data has become a key problem. However, the use of a single parameter (e.g. the box dimension of the graph) is too reductive as a classification tool in many situations that are met in applications. This explains the success of multifractal analysis, which is a way to associate a whole collection of fractal-based parameters to a function. Its purpose is twofold: on the mathematical side, it allows one to determine the size of the sets of points where a function has a given Hölder exponent; on the signal processing side, it yields new collections of parameters associated to the considered signal and which can be used for classification, model selection, or for parameter selection inside a parametric setting. The main advances in the subject came from a better understanding of the interactions between these two motivations, e.g., see [3] and references therein for recent review papers.

Despite the fact that multifractal analysis has traditionally been based on the Hölder exponent, it is not the only characterization of pointwise regularity that can be used. Therefore, our goal in the present contribution is to analyze alternative pointwise exponents and the information they provide.

In Sect. 2 we review the possible pointwise exponents of functions, and explain in which context each can be used.

In Sect. 3 we focus on the p-exponent, derive some of its properties, and investigate what information it yields concerning the lacunarity of the local behavior of the function near a singularity.

In Sect. 4 we recall the derivation of the multifractal formalism and give applications to a simple model of a random process which displays multifractal behavior: Lacunary wavelet series.

We conclude with remarks on the relationship between the existence of p-exponents and the sparsity of the wavelet expansion.

This paper partly reviews elements on the p-exponent which are scattered in the literature, see e.g. [2, 8, 15, 16, 21]. New material starts with the introduction and analysis of the lacunarity exponent in Sect. 2.3, the analysis of thin chirps in Sect. 3.5, and all following sections, except for the brief reminder on the multifractal formalism in Sect. 4.1.

2 Pointwise Exponents

In this section, unless otherwise specified, we assume that $f \in L^1_{loc}(\mathbb{R})$. An important remark concerning the definition of pointwise Hölder regularity is that if (1.2) holds (even for $\gamma < 0$), then f is bounded in any annulus $0 < r \leq |x - x_0| \leq R$. It follows that, if an estimate such as (1.2) holds for all x_0, then f will be locally bounded, except perhaps at isolated points. For this reason, one usually assumes that the considered function f is (everywhere) locally bounded. It follows that (1.2) holds for $\gamma = 0$ so that the Hölder exponent is always nonnegative.

2.1 Uniform Hölder Regularity

An important issue therefore is to determine if the regularity assumption $f \in L^\infty_{loc}$ is satisfied for real life data. This can be done in practice by first determining their **uniform Hölder exponent**, which is defined as follows.

Recall that Lipschitz spaces $C^s(\mathbb{R})$ are defined for $0 < s < 1$ by

$$f \in L^\infty \quad \text{and} \quad \exists C, \quad \forall x, y, \qquad |f(x) - f(y)| \le C|x-y|^s.$$

If $s > 1$, they are then defined by recursion on $[s]$ by the condition: $f \in C^s(\mathbb{R})$ if $f \in L^\infty$ and if its derivative f' (taken in the sense of distributions) belongs to $C^{s-1}(\mathbb{R})$. If $s < 0$, then the C^s spaces are composed of distributions, also defined by recursion on $[s]$ as follows: $f \in C^s(\mathbb{R})$ if f is a derivative (in the sense of distributions) of a function $g \in C^{s+1}(\mathbb{R})$. We thus obtain a definition of the C^s spaces for any $s \notin \mathbb{Z}$ (see [22] for $s \in \mathbb{Z}$, which we will however not need to consider in the following). A distribution f belongs to C^s_{loc} if $f\varphi \in C^s$ for every C^∞ compactly supported function φ.

Definition 2.1 The uniform Hölder exponent of a tempered distribution f is

$$H_f^{min} = \sup\{s : f \in C^s_{loc}(\mathbb{R})\}. \tag{2.1}$$

This definition does not make any a priori assumption on f: The uniform Hölder exponent is defined for any tempered distribution, and it can be positive or negative. More precisely:

- If $H_f^{min} > 0$, then f is a locally bounded function,
- if $H_f^{min} < 0$, then f is not a locally bounded function.

In practice, this exponent is determined through the help of the wavelet coefficients of f. By definition, an **orthonormal wavelet basis** is generated by a couple of functions $(\varphi, \ \psi)$, which, in our case, will either be in the Schwartz class, or smooth and compactly supported (in that case, wavelets are assumed to be smoother than the regularity exponent of the considered space). The functions $\varphi(x-k), \quad k \in \mathbb{Z}$, together with $2^{j/2}\psi(2^j x - k), \quad j \ge 0, \ k \in \mathbb{Z}$, form an orthonormal basis of $L^2(\mathbb{R})$. Thus any function $f \in L^2(\mathbb{R})$ can be written

$$f(x) = \sum_k c_k \, \varphi(x-k) + \sum_{j \ge 0} \sum_{k \in \mathbb{Z}} c_{j,k} \, \psi(2^j x - k),$$

where the wavelet coefficients of f are given by

$$c_k = \int \varphi(t-k) f(t) dt \quad \text{and} \quad c_{j,k} = 2^j \int \psi(2^j t - k) f(t) dt. \tag{2.2}$$

An important remark is that these formulas also hold in many different functional settings (such as the Besov or Sobolev spaces of positive or negative regularity), provided that the picked wavelets are smooth enough (and that the integrals (2.2) are understood as duality products).

Instead of using the indices (j,k), we will often use dyadic intervals: Let

$$\lambda\ (=\lambda(j,k))\ = \left[\frac{k}{2^j}, \frac{k+1}{2^j}\right) \tag{2.3}$$

and, accordingly: $c_\lambda = c_{j,k}$ and $\psi_\lambda(x) = \psi(2^j x - k)$. Indexing by dyadic intervals will be useful in the sequel because the interval λ indicates the localization of the corresponding wavelet: When the wavelets are compactly supported, then, $\exists C > 0$ such that when $supp(\psi) \subset [-C/2, C/2]$, then $supp(\psi_\lambda) \subset 2C\lambda$.

In practice, H_f^{min} can be derived directly from the wavelet coefficients of f through a simple regression in a log-log plot; indeed, it follows from the wavelet characterization of the spaces C^s, see [22], that:

$$H_f^{min} = \liminf_{j\to+\infty} \quad \frac{\log\left(\sup_k |c_{j,k}|\right)}{\log(2^{-j})}. \tag{2.4}$$

This estimation procedure has been studied in more detail in [20]. Three examples of its numerical application to real-world functions are provided in Fig. 1.

A multifractal analysis based on the Hölder exponent can only be performed if f is locally bounded. A way to determine if this is the case consists in first

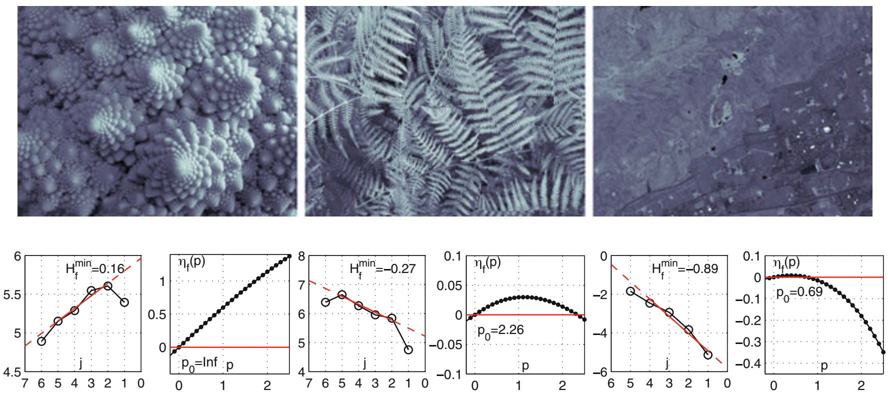


Fig. 1 **Real-world images** (*top row*) of Romanesco broccoli (*left column*), fern leaves (*center column*) and a patch of a hyperspectral image of the Moffett field, acquired by the AVIRIS instrument (spectral band 90, *right column*). *Bottom row*: corresponding numerical estimation of uniform Hölder exponents H_f^{min}, wavelet scaling functions $\eta_f(p)$ and critical Lebesgue indices p_0, respectively

checking if $H_f^{min} > 0$. This quantity is perfectly well-defined for mathematical functions or stochastic processes; e.g. for Brownian motion, $H_f^{min} = 1/2$, and for Gaussian white noise, $H_f^{min} = -1/2$. However the situation may seem less clear for experimental signals; indeed any data acquisition device yields a finite set of locally averaged quantities, and one may argue that such a finite collection of data (which, by construction, is bounded) can indeed be modeled by a locally bounded function. This argument can only be turned by revisiting the way that (2.4) is computed in practice: Estimation is performed through a linear regression in log-log coordinates **on the range of scales available in the data** and H_f^{min} can indeed be found negative for a finite collection of data. At the modeling level, this means that a mathematical model which would display the same linear behavior in log-log coordinates **at all scales** would satisfy $H_f^{min} < 0$.

The quantity H_f^{min} can be found either positive or negative depending on the nature of the application. For instance, velocity turbulence data and price time series in finance are found to always have $H_f^{min} > 0$, while aggregated count Internet traffic time series always have $H_f^{min} < 0$. For biomedical applications (cf. e.g., fetal heart rate variability) as well as for image processing, H_f^{min} can commonly be found either positive or negative (see Fig. 1) [1, 3, 19, 20, 28]. This raises the problem of using other pointwise regularity exponents that would not require the assumption that the data are locally bounded. We now introduce such exponents.

2.2 The p-Exponent for p ≥ 1

The introduction of p-exponents is motivated by the necessity of introducing regularity exponents that could be defined even when H_f^{min} is found to be negative; $T_\alpha^p(x_0)$ regularity, introduced by A. Calderón and A. Zygmund in [8], has the advantage of only making the assumption that f locally belongs to $L^p(\mathbb{R})$.

Definition 2.2 Let $p \geq 1$ and assume that $f \in L^p_{loc}(\mathbb{R})$. Let $\alpha \in \mathbb{R}$; the function f belongs to $T_\alpha^p(x_0)$ if there exists C and a polynomial P_{x_0} of degree less than α such that, for r small enough,

$$\left(\frac{1}{2r}\int_{x_0-r}^{x_0+r} |f(x) - P_{x_0}(x)|^p dx\right)^{1/p} \leq Cr^\alpha. \tag{2.5}$$

Note that the **Taylor polynomial** P_{x_0} of f at x_0 might depend on p. However, one can check that only its degree does (because the best possible α that one can pick in (2.5) depends on p so that its integer part may vary with p, see [2]). Therefore we introduce no such dependency in the notation, which will lead to no ambiguity afterwards.

The ***p*-exponent** of f at x_0 is defined as

$$h_f^p(x_0) = \sup\{\alpha : f \in T_\alpha^p(x_0)\}. \tag{2.6}$$

The condition that f locally belongs to $L^p(\mathbb{R})$ implies that (2.5) holds for $\alpha = -1/p$, so that $h_f^p(x_0) \geq -1/p$.

We will consider in the following "archetypical" pointwise singularities, which are simple toy-examples of singularities with a specific behavior at a point. They will illustrate the new notions we consider and they will also supply benchmarks on which we can compute exactly what these new notions allow us to quantify. These toy-examples will be a test for the adequacy between these mathematical notions and the intuitive behavior that we expect to quantify. The first (and most simple) "archetypical" pointwise singularities are the **cusp singularities**.

Let $\alpha \in \mathbb{R} - 2\mathbb{N}$ be such that $\alpha > -1$. The **cusp of order α at** 0 is the function

$$\mathcal{C}_\alpha(x) = |x|^\alpha. \tag{2.7}$$

The case $\alpha \in 2\mathbb{N}$ is excluded because it leads to a C^∞ function. However, if $\alpha = 2n$, one can pick

$$\mathcal{C}_{2n}(x) = x|x|^{2n-1},$$

in order to cover this case also.

If $\alpha \geq 0$, then the cusp $\mathcal{C}_\alpha$ is locally bounded and its Hölder exponent at 0 is well-defined and takes the value α. If $\alpha > -1/p$, then its p-exponent at 0 is well-defined and also takes the value α, as in the Hölder case. (Condition $\alpha > -1/p$ is necessary and sufficient to ensure that $\mathcal{C}_\alpha$ locally belongs to L^p.) Examples for cusps with several different values of α are plotted in Fig. 2.

If $f \in L^p_{loc}$ in a neighborhood of x_0 for a $p \geq 1$, let us define the **critical Lebesgue index** of f at x_0 by

$$p_0(f) = \sup\{p : f \in L^p_{loc}(\mathbb{R}) \text{ in a neighborhood of } x_0\}. \tag{2.8}$$

The importance of this exponent comes from the fact that it tells in practice for which values of p a p-exponent based multifractal analysis can be performed. Therefore, its numerical determination is an important prerequisite that should not be bypassed in applications. In Sect. 3.1 we will extend the definition of $p_0(f)$ to situations where $f \notin L^1_{loc}$ and show how it can be derived from another quantity, the wavelet scaling function, which can be effectively computed on real-life data.

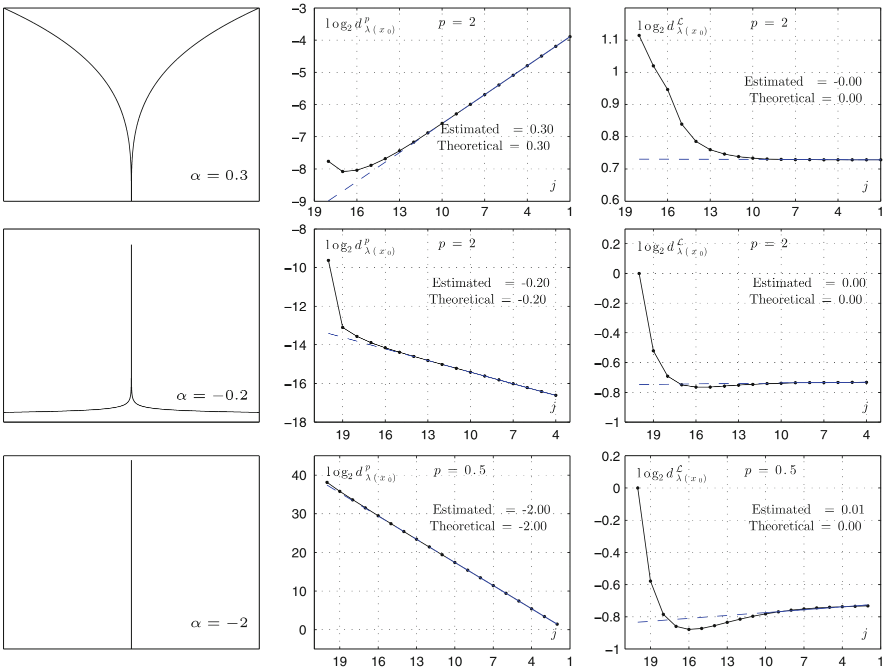


Fig. 2 **Cusps** with exponents $\alpha = \{+0.3, -0.2, -2\}$ (from *top* to *bottom row*, respectively): functions (*left column*) and estimation of p-exponents and lacunarity exponents (*center* and *right column*, respectively). The critical Lebesgue indices are given by $p_0 = \{+\infty, 5, 0.5\}$, respectively

2.3 The Lacunarity Exponent

The p-exponent at x_0 is defined on the interval $[1, p_0(f)]$ or $[1, p_0(f))$; when the p-exponent does not depend on p on this interval, we will say that f has a ***p*-invariant singularity** at x_0. Thus, cusps are p-invariant singularities.

This first example raises the following question: Is the notion of p-exponent only relevant as an extension of the Hölder exponent to non-locally bounded functions? Or can it take different values with p, even for bounded functions? And, if such is the case, how can one characterize the additional information thus supplied? In order to answer this question, we introduce a second type of archetypical singularities, the **lacunary singularities**, which will show that the p-exponent may be non-constant. We first need to recall the geometrical notion of **accessibility exponent** which quantifies the **lacunarity** of a set at a point, see [17]. We denote by $\mathcal{M}(A)$ the Lebesgue measure of a set A.

Definition 2.3 Let $\Omega \subset \mathbb{R}$. A point x_0 of the boundary of Ω is α-accessible if there exist $C > 0$ and $r_0 > 0$ such that $\forall r \leq r_0$,

$$\mathcal{M}\,(\Omega \cap B(x_0, r)) \leq C r^{\alpha+1}. \tag{2.9}$$

The supremum of all values of α such that (2.9) holds is called the accessibility exponent of Ω at x_0. We will denote it by $\mathcal{E}_{x_0}(\Omega)$.

Note that $\mathcal{E}_{x_0}(\Omega)$ is always nonnegative. If it is strictly positive, then Ω is lacunary at x_0. The accessibility exponent supplies a way to estimate, through a log-log plot regression, the "size" of the part of Ω which is contained in arbitrarily small neighborhoods of x_0. The following sets illustrate this notion.

Let ω and γ be such that $0 < \gamma \leq \omega$; the set $U_{\omega,\gamma}$ is defined as follows. Let

$$I^j_{\omega,\gamma} = [2^{-\omega j}, 2^{-\omega j} + 2^{-\gamma j}]; \quad \text{then} \quad U_{\omega,\gamma} = \bigcup_{j \geq 0} I^j_{\omega,\gamma}. \tag{2.10}$$

Clearly, at the origin,

$$\mathcal{E}_0(U_{\omega,\gamma}) = \frac{\gamma}{\omega} - 1. \tag{2.11}$$

We now construct univariate functions $F_{\alpha,\gamma} : \mathbb{R} \to \mathbb{R}$ which permit us to better understand the conditions under which p-exponents will differ. These functions will have a lacunary support in the sense of Definition 2.3.

Let ψ be the Haar wavelet: $\psi = 1_{[0,1/2)} - 1_{[1/2,1)}$ and

$$\theta(x) = \psi(2x) - \psi(2x - 1)$$

(so that θ has the same support as ψ but its two first moments vanish).

Definition 2.4 Let $\alpha \in \mathbb{R}$ and $\gamma > 1$. The lacunary comb $F^{\alpha}_{\omega,\gamma}$ is the function

$$F^{\alpha}_{\omega,\gamma}(x) = \sum_{j=1}^{\infty} 2^{-\alpha j} \theta\left(2^{\gamma j}(x - 2^{-\omega j})\right). \tag{2.12}$$

Note that its singularity is at $x_0 = 0$. Numerical examples of lacunary combs are provided in Fig. 3.

Note that the support of $F^{\alpha}_{\omega,\gamma}$ is $U_{\omega,\gamma}$ so that the accessibility exponent at 0 of this support is given by (2.11). The function $F^{\alpha}_{\omega,\gamma}$ is locally bounded if and only if $\alpha \geq 0$. Assume that $\alpha < 0$; then $F^{\alpha}_{\omega,\gamma}$ locally belongs to L^p if and only if $\alpha > -\gamma/p$. When such is the case, a straightforward computation yields that its p-exponent at 0

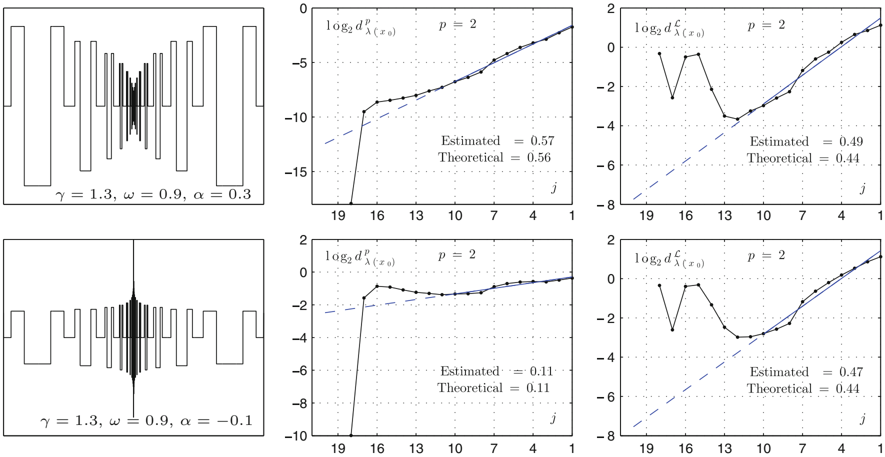


Fig. 3 **Lacunary combs** with $p_0 = +\infty$ (*top row*) and $p_0 = 9.1$ (*bottom row*): functions (*left column*) and estimation of p-exponents and lacunarity exponents (*center* and *right column*, respectively)

is given by

$$h^p_{F^{\alpha}_{\omega,\gamma}}(x_0) = \frac{\alpha}{\omega} + \left(\frac{\gamma}{\omega} - 1\right)\frac{1}{p}. \tag{2.13}$$

In contradistinction with the cusp case, the p-exponent of $F^{\alpha}_{\omega,\gamma}$ at 0 is not a constant function of p. Let us see how the variations of the mapping $p \to h^p_f(x_0)$ are related with the lacunarity of the support of f, in the particular case of $F^{\alpha}_{\omega,\gamma}$. We note that this mapping is an affine function of the variable $q = 1/p$ (which, in this context, is a more natural parameter than p) and that the accessibility exponent of the support of $F^{\alpha}_{\omega,\gamma}$ can be recovered by a derivative of this mapping with respect to q. The next question is to determine the value of q at which this derivative should be taken. This toy-example is too simple to give a clue since any value of q would lead to the same value for the derivative. We want to find if there is a more natural one, which would lead to a *canonical* definition for the lacunarity exponent. It is possible to settle this point through the following simple perturbation argument: Consider a new singularity F that would be the sum of two functions $F_1 = F^{\alpha_1}_{\omega_1,\gamma_1}$ and $F_2 = F^{\alpha_2}_{\omega_2,\gamma_2}$ with

$$0 < \alpha_1 < \alpha_2 \quad \text{and} \quad \gamma_1 > \gamma_2. \tag{2.14}$$

The p-exponent of F (now expressed in the q variable, where $q = 1/p$) is given by

$$q \mapsto h^{\frac{1}{q}}_f(x_0) = \min\left[\frac{\alpha_1}{\omega} + \left(\frac{\gamma_1}{\omega} - 1\right)q\,,\ \ \frac{\alpha_2}{\omega} + \left(\frac{\gamma_2}{\omega} - 1\right)q\right]. \tag{2.15}$$

The formula for the lacunarity exponent should yield the lacunarity of the most irregular component of F; since $F \in L^{\infty}_{loc}$, the Hölder exponent is the natural way to measure this irregularity. In this respect, the most irregular component is F_1; the lacunarity exponent should thus take the value $\left(\frac{\gamma_1}{\omega} - 1\right)$. But, since (2.14) allows the shift in slope of the function (2.15) from $\left(\frac{\gamma_1}{\omega} - 1\right)$ to $\left(\frac{\gamma_2}{\omega} - 1\right)$ to take place at a q arbitrarily close to 0, the only way to obtain this desired result in any case is to pick the derivative of the mapping $q \to h_f^{1/q}(x_0)$ precisely at $q = 0$.

A similar perturbation argument can be developed if $p_0(f) < \infty$ with the conclusion that the derivative should be estimated at the smallest possible value of q, i.e. for

$$q = q_0(f) := \frac{1}{p_0(f)};$$

hence the following definition of the **lacunarity exponent**.

Definition 2.5 Let $f \in L^p_{loc}$ in a neighborhood of x_0 for a $p > 1$, and assume that the p-exponent of f is finite in a left neighborhood of $p_0(f)$. The lacunarity exponent of f at x_0 is

$$\mathcal{L}_f(x_0) = \frac{\partial}{\partial q}\left(h_f^{1/q}(x_0)\right)_{q=q_0(f)^+}. \tag{2.16}$$

Remarks

- Even if the p-exponent is not defined at $p_0(f)$, nonetheless, because of the concavity of the mapping $q \to h_f^{1/q}(x_0)$ (see Proposition 3.2 below), its right derivative is always well-defined, possibly as a limit.
- As expected, the lacunarity exponent of a cusp vanishes, whereas the lacunarity exponent of a lacunary comb coincides with the accessibility exponent of its support.
- The condition $\mathcal{L}_f(x_0) \neq 0$ does not mean that the support of f (or of $f - P$) has a positive accessibility exponent (think of the function $F^{\alpha}_{\omega,\gamma} + g$ where g is a C^{∞} but nowhere polynomial function).
- The definition supplied by (2.16) bears similarity with the definition of the oscillation exponent (see [4, 20] and ref. therein) which is also defined through a derivative of a pointwise exponent; but the variable with respect to which the derivative is computed is the order of a fractional integration. The relationships between these two exponents will be investigated in a forthcoming paper [21].

3 Properties of the p-Exponent

In signal and image processing, one often meets data that cannot be modeled by functions $f \in L^1_{loc}$, see Fig. 1. It is therefore necessary to set the analysis in a wider functional setting, and therefore to extend the notion of $T^p_\alpha(x_0)$ regularity to the case $p < 1$.

3.1 The Case $p < 1$

The standard way to perform this extension is to consider exponents in the setting of the real Hardy spaces H^p (with $p < 1$) instead of L^p spaces, see [15, 16]. First, we need to extend the definitions that we gave to the range $p \in (0, 1]$. The simplest way is to start with the wavelet characterization of L^p spaces, which we now recall.

We denote indifferently by $\chi_{j,k}$ or χ_λ the characteristic function of the interval λ ($= \lambda_{j,k}$) defined by (2.3). The **wavelet square function** of f is

$$\mathcal{W}_f(x) = \left(\sum_{(j,k)\in\mathbb{Z}^2} |c_{j,k}|^2 \chi_{j,k}(x) \right)^{1/2}.$$

Then, for $p > 1$,

$$f \in L^p(\mathbb{R}) \iff \int_{\mathbb{R}} \left(\mathcal{W}_f(x)\right)^p dx < \infty, \tag{3.1}$$

see [22]. The quantity $\left(\int \left(\mathcal{W}_f(x)\right)^p dx\right)^{1/p}$ is thus equivalent to $\| f \|_p$. One can then take the characterization supplied by (3.1) when $p > 1$ as a definition of the Hardy space H^p (when $p \leq 1$); note that this definition yields equivalent quantities when the (smooth enough) wavelet basis is changed, see [22]. This justifies the fact that we will often denote by L^p the space H^p, which will lead to no confusion; indeed, when $p \leq 1$ this notation will refer to H^p, and, when $p > 1$ it will refer to L^p.

Note that, if $p = 1$, (3.1) does not characterize the space L^1 but a strict subspace of L^1 (the real Hardy space H^1, which consists of functions of L^1 whose Hilbert transform also belongs to L^1, see [22]).

Most results proved for the L^p setting will extend without modification to the H^p setting. In particular, T^p_α regularity can be extended to the case $p \leq 1$ and has the same wavelet characterization, see [13]. All definitions introduced previously therefore extend to this setting.

The definition of $T^p_\alpha(x_0)$ regularity given by (2.5) is a size estimate of an L^p norm restricted to intervals $[x_0 - r, x_0 + r]$. Since the elements of H^p can be distributions, the restriction of f to an interval cannot be done directly (multiplying a distribution by a non-smooth function, such as a characteristic function, does not always make sense). This problem can be solved as follows: If I is an open interval, one defines $\| f \|_{H^p(I)} = \inf \| g \|_p$, where the infimum is taken on the $g \in H^p$ such that $f = g$ on I. The T^p_α condition for $p \leq 1$ is then defined by:

$$f \in T^p_\alpha(x_0) \quad \Longleftrightarrow \quad \| f \|_{H^p((x_0-r,x_0+r))} \leq C\, r^{\alpha+1/p},$$

also when $p < 1$. We will show below that the p-exponent takes values in $[-1/p, +\infty]$.

3.2 When Can One Use p-Exponents?

We already mentioned that, in order to use the Hölder exponent as a way to measure pointwise regularity, we need to check that the data are locally bounded, a condition which is implied by the criterion $H_f^{min} > 0$, which is therefore used as a practical prerequisite. Similarly, in order to use a p-exponent based multifractal analysis, we need to check that the data locally belong to L^p or H^p, a condition which can be verified in practice through the computation of the *wavelet scaling function*, which we now recall.

The Sobolev space $L^{p,s}$ is defined by

$$\forall s \in \mathbb{R},\ \forall p > 0, \qquad f \in L^{p,s} \quad \Longleftrightarrow \quad (Id - \Delta)^{s/2} f \in L^p,$$

where the operator $(Id - \Delta)^{s/2}$ is the Fourier multiplier by $(1 + |\xi|^2)^{s/2}$, and we recall our convention that L^p denotes the space H^p when $p \leq 1$, so that Sobolev spaces are defined also for $p \leq 1$.

Definition 3.1 Let f be a tempered distribution. The wavelet scaling function of f is defined by

$$\forall p > 0, \qquad \eta_f(p) = p\ \sup\{s : f \in L^{p,s}\}. \tag{3.2}$$

Thus, $\forall p > 0$:

- If $\eta_f(p) > 0$ then $f \in L^p_{loc}$.
- If $\eta_f(p) < 0$ then $f \notin L^p_{loc}$.

The wavelet characterization of Sobolev spaces implies that the wavelet scaling function can be expressed as (cf. [11])

$$\forall p > 0, \qquad \eta_f(p) = \liminf_{j \to +\infty} \frac{\log\left(2^{-j} \displaystyle\sum_k |c_{j,k}|^p\right)}{\log(2^{-j})}. \tag{3.3}$$

This provides a practical criterion for determining if data locally belong to L^p, supplied by the condition $\eta_f(p) > 0$. The following bounds for $p_0(f)$ follow:

$$\sup\{p : \eta_f(p) > 0\} \leq p_0(f) \leq \inf\{p : \eta_f(p) < 0\},$$

which (except in the very particular cases where η_f vanishes identically on an interval) yields the exact value of $p_0(f)$.

In applications, data with very different values of $p_0(f)$ show up; therefore, in practice, the mathematical framework supplied by the whole range of p is relevant. As an illustration, three examples of real-world images with positive and negative uniform Hölder exponents and with critical Lebesgue indices above and below $p_0 = 1$ are analyzed in Fig. 1.

3.3 Wavelet Characterization of p-Exponents

In order to compute and prove properties of p-exponents we will need the exact wavelet characterization of $T^p_\alpha(x_0)$, see [13, 15]. Let λ be a dyadic interval; 3λ will denote the interval of same center and three times wider (it is the union of λ and its two closest neighbors). For $x_0 \in \mathbb{R}^d$, denote by $\lambda_j(x_0)$ the dyadic cube of width 2^{-j} which contains x_0. The **local square functions** at x_0 are the sequences defined for $j \geq 0$ by

$$\mathcal{W}^j_{f,x_0}(x) = \left(\sum_{\lambda \subset 3\lambda_j(x_0)} |\, c_\lambda|^2 \chi_\lambda(x) \right)^{1/2}.$$

Recall that (cf. [13])

$$f \in T^p_\alpha(x_0) \quad \text{if and only if } \exists C > 0,\ \forall j \geq 0 \qquad \left\| \mathcal{W}^j_{f,x_0} \right\|_p \leq C\, 2^{-(\alpha+1/p)j}. \tag{3.4}$$

The following result is required for the definition of the lacunarity exponent in (2.16) to make sense, and implies that Definition 2.5 also makes sense when $p_0(f) < 1$.

Proposition 3.2 *Let $p, q \in (0, +\infty]$, and suppose that $f \in T^p_\alpha(x_0) \cap T^q_\beta(x_0)$; let $\theta \in [0, 1]$. Then $f \in T^r_\gamma(x_0)$, where*

$$\frac{1}{r} = \frac{\theta}{p} + \frac{1-\theta}{q} \quad \textit{and} \quad \gamma = \theta\alpha + (1-\theta)\beta.$$

It follows that the mapping $q \to h^{1/q}_f(x_0)$ is concave on its domain of definition.

Proof When $p, q < \infty$, the result is a consequence of (3.4). Hölder's inequality implies that

$$\left\| \mathcal{W}^j_{f,x_0} \right\|_r \leq \left\| \mathcal{W}^j_{f,x_0} \right\|_p^{\theta/p} \left\| \mathcal{W}^j_{f,x_0} \right\|_q^{(1-\theta)/q}.$$

We thus obtain the result for $p, q < \infty$. The case when p or $q = +\infty$ does not follow, because there exists no exact wavelet characterization of $C^\alpha(x_0) = T^\infty_\alpha(x_0)$; however, when $p, q > 1$, one can use the initial definition of $T^p_\alpha(x_0)$ and $C^\alpha(x_0)$ through local L^p and L^∞ norms and the result also follows from Hölder's inequality; hence Proposition 3.2 holds.

If $f \in H^p$, then $\| \mathcal{W}_f \|_p \leq C$. Since $\mathcal{W}^j_f \leq \mathcal{W}_f$, it follows that $\| \mathcal{W}^j_f \|_p \leq C$, so that (3.4) holds with $\alpha = -1/p$. Thus p-exponents are always larger than $-1/p$ (which extends to the range $p < 1$ the result already mentioned for $p \geq 1$). Note that this bound is compatible with the existence of singularities of arbitrary large negative order (by picking p close to 0). The example of cusps will now show that the p-exponent can indeed take values down to $-1/p$.

3.4 Computation of p-Exponents for Cusps

Typical examples of distributions for which the p-exponent is constant (see Proposition 3.3 below) and equal to a given value $\alpha < -1$ are supplied by the cusps $\mathcal{C}_\alpha$, whose definition can be extended to the range $\alpha \leq -1$ as follows: First, note that cusps cannot be defined directly for $\alpha \leq -1$ by (2.7) because they do not belong to L^1_{loc} so that they would be ill-defined even in the setting of distributions (their integral against a C^∞ compactly supported function φ may diverge). Instead, we use the fact that, if $\alpha > 1$, then $\mathcal{C}''_\alpha = \alpha(\alpha-1)\mathcal{C}_{\alpha-2}$, which indicates a way to define by recursion the cusps $\mathcal{C}_\alpha$, when $\alpha < -1$ and $\alpha \notin \mathbb{Z}$, as follows:

$$\text{if} \quad \alpha < 0, \qquad \mathcal{C}_\alpha = \frac{1}{(\alpha+1)(\alpha+2)} \mathcal{C}''_{\alpha+2},$$

where the derivative is taken in the sense of distributions. The $\mathcal{C}_\alpha$ are thus defined as distributions when α is not a negative integer. It can also be done when α is a negative integer, using the following definition for $\alpha = 0$ and -1:

$$\mathcal{C}_0 = \log(|x|) \qquad \text{and} \qquad \mathcal{C}_{-1} = \mathcal{C}'_0 = P.V.\left(\frac{1}{x}\right),$$

where P.V. stands for "principal value".

Proposition 3.3 *If $\alpha \geq 0$, the cusp $\mathcal{C}_\alpha$ belongs to L^∞_{loc} and its p-exponent is α. If $\alpha < 0$, the cusp $\mathcal{C}_\alpha$ belongs to L^p_{loc} for $p < -1/\alpha$ and its p-exponent is α.*

Proof of Proposition 3.3 The case $\alpha \geq 0$ and $p \geq 1$ has already been considered in [17, 21]. In this case, the computation of the p-exponent is straightforward. Note that, when $\alpha \in (-1, 0)$ and $p \geq 1$ the computations are similar. We thus focus on the distribution case, i.e. when $p < 1$. The global and pointwise regularity will be determined through an estimation of the wavelet coefficients of the cusp. We use a smooth enough, compactly supported wavelet basis and we denote by $c_{j,k}$ the wavelet coefficients of the cusp

$$c_{j,k} = 2^j \langle \psi_{j,k} | \mathcal{C}_\alpha \rangle.$$

The selfsimilarity of the cusp implies that

$$\forall j, k \qquad c_{j,k} = 2^{-\alpha j} c_{0,k}; \tag{3.5}$$

additionally, as soon as k is large enough so that the support of $\psi(x-k)$ does not intersect the origin, the cusp is C^∞ in the support of $\psi(x-k)$ and coincides with the function $|x|^\alpha$. An integration by parts then yields that, for any N smaller than the global regularity of the wavelet,

$$c_{0,k} = (-1)^N \int \psi^{(-N)}(x-k)\, \alpha(\alpha-1)\cdots(\alpha-N)|x|^{\alpha-N} dx,$$

so that the sequence $c_{0,k}$ satisfies

$$|c_{0,k}| \leq \frac{C_N}{(1+|k|)^N} \tag{3.6}$$

where N can be picked arbitrarily large. The estimation of the L^p norm of the wavelet square function follows easily from (3.5) and (3.6), and so does the lower bound for the p-exponent. The upper bound is obtained by noticing that one of the $c_{0,k}$ does not vanish (otherwise, all $c_{j,k}$ would vanish, and the cusp would be a smooth function at the origin). Therefore, there exists at least one k_0 such that $\forall j$, $c_{j,k_0} = C 2^{-\alpha j}$, and the wavelet characterization of T^p_α regularity then yields that $h^p(x_0) \leq \alpha$.

Three examples of cusps and numerical estimates of their p-exponents and lacunarity exponents are plotted in Fig. 2.

3.5 Wavelet Characterization and Thin Chirps

In practice, we will derive T^p_α regularity from simpler quantities than the local square functions. The ***p*-leaders** of f are defined by local l^p norms of wavelet coefficients as follows:

$$d_\lambda^p = \left(\sum_{\lambda' \subset 3\lambda} |c_{\lambda'}|^p 2^{-(j'-j)} \right)^{1/p} \tag{3.7}$$

(they are finite if $f \in L^p_{loc}(\mathbb{R}^d)$, see [17]). Note that, if $p = +\infty$, the corresponding quantity d_λ^∞ is usually denoted by d_λ and simply called the **wavelet leaders**; we have

$$d_\lambda := d_\lambda^\infty = \sup_{\lambda' \subset 3\lambda} |c_{\lambda'}|. \tag{3.8}$$

The notion of T^p_α regularity can be related to p-leader coefficients (see [16, 17, 20]):

$$\text{If } \eta_f(p) > 0, \text{ then} \quad h_f^p(x_0) = \liminf_{j \to +\infty} \frac{\log\left(d^p_{\lambda_j(x_0)}\right)}{\log(2^{-j})}. \tag{3.9}$$

Our purpose in this section is to introduce new "archetypical" pointwise singularities which will yield examples where the p-exponent and the lacunarity exponent can take arbitrary values. Because of (3.9), it is easier to work with examples that are defined directly by their wavelet coefficients on a smooth wavelet basis. We therefore develop new examples rather than extending the lacunary combs of Sect. 2.3.

Definition 3.4 Let $a, b \in (0, 1)$ satisfying $0 < b < 1 - a$, and let $\alpha \in \mathbb{R}$. The thin chirp $\mathcal{T}_{a,b,\alpha}$ is defined by its wavelet series

$$\mathcal{T}_{a,b,\alpha} = \sum_{j \geq 0} \sum_{k \in \mathbb{Z}} c_{j,k} \; \psi_{j,k},$$

where

$$\begin{aligned} c_{j,k} &= 2^{-\alpha j} \quad \text{if} \quad k \in [2^{(1-a)j}, 2^{(1-a)j} + 2^{bj}] \\ &= 0 \qquad\quad \text{otherwise.} \end{aligned}$$

The following results are straightforward, using the wavelet characterization of L^p and T^p_α regularity.

Proposition 3.5 *The thin chirp $\mathcal{T}_{a,b,\alpha}$ is bounded if and only if $\alpha > 0$.*

$$\textit{If} \quad \alpha \leq 0, \qquad p_0(\mathcal{T}_{a,b,\alpha}) = \frac{1-b}{-\alpha}.$$

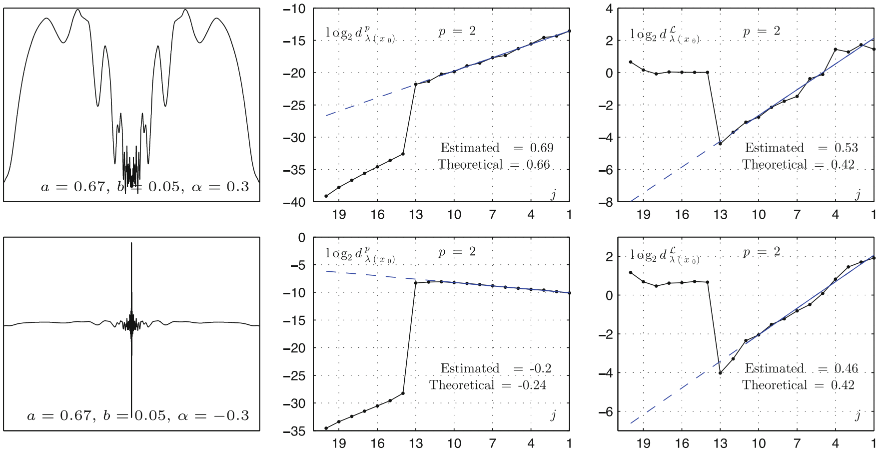


Fig. 4 **Thin chirps** with $p_0 = \infty$ (*top row*) and $p_0 = 3.2$ (*bottom row*): functions (*left column*) and estimation of p-exponents and lacunarity exponents (*center* and *right column*, respectively)

The p-exponent of $\mathcal{T}_{a,b,\alpha}$ at the origin is

$$h^p_{\mathcal{T}_{a,b,\alpha}}(0) = \frac{1-a-b}{a} q + \frac{\alpha}{a}.$$

Note that, if the wavelets are compactly supported, then for j large enough the pack of 2^{bj} successive wavelets with non-vanishing coefficients covers an interval of length $2^{-j}2^{bj}$ at a distance 2^{-aj} from the origin, so that the accessibility exponent of the support of $\mathcal{T}_{a,b,\alpha}$ is $(1-a-b)/a$: Thus, it coincides with the lacunarity exponent of $\mathcal{T}_{a,b,\alpha}$ as expected.

Illustrations of thin chirps and the numerical estimation of their p-exponents and lacunarity exponents are provided in Fig. 4.

3.6 p-Exponent Analysis of Measures

Several types of measures (such as multiplicative cascades) played a central role in the development of multifractal analysis. Since measures (usually) are not L^1 functions, their p-exponent for $p \geq 1$ is not defined. Therefore, it is natural to wonder if it can be the case when $p < 1$. This is one of the purposes of Proposition 3.6, which yields sufficient conditions under which a measure μ satisfies $\eta_\mu(p) > 0$ for $p < 1$, which will imply that its p-exponent multifractal analysis can be performed. An important by-product of using p-exponents for $p \leq 1$ is that it offers a common setting to treat pointwise regularity of measures and functions.

Recall that $\overline{\dim}_B(A)$ denotes the upper box dimension of the set A.

Proposition 3.6 *Let μ be a measure; then its wavelet scaling function satisfies $\eta_\mu(1) \geq 0$. Furthermore, if μ does not have a density which is an L^1 function, then $\eta_\mu(1) = 0$.*

Additionally, if μ is a singular measure whose support supp(μ) satisfies

$$\delta_\mu := \overline{\dim}_B(supp(\mu)) < 1, \tag{3.10}$$

then

$$\forall p < 1, \qquad \eta_\mu(p) \geq (1-\delta_\mu)(1-p), \tag{3.11}$$

and

$$\forall p > 1, \qquad \eta_\mu(p) \leq (1-\delta_\mu)(1-p). \tag{3.12}$$

Remarks

- (3.11) expresses the fact that, if μ has a small support, then its Sobolev regularity is increased for $p < 1$. This is somehow counterintuitive, since one expects a measure to become more singular when the size of its support shrinks; on the other hand (3.12) expresses that this is actually the case when $p > 1$.
- Condition $\delta_\mu < 1$ is satisfied if μ is supported by a Cantor-like set, or by a selfsimilar set satisfying Hutchinson's open set condition.
- (3.11) has an important consequence for the multifractal analysis of measures: Indeed, if $\delta_\mu < 1$, then $\eta_\mu(p) > 0$ for $p < 1$, so that the classical mathematical results concerning the multifractal analysis based on the p-exponent apply, see Sect. 4.
- A slightly different problem was addressed by H. Triebel: In [27], he determined under which conditions the scaling functions commonly used in the multifractal analysis of probability measures (see (4.4) below) can be recovered through Besov or Triebel-Lizorkin norms (or semi-norms).

Proof of Proposition 3.6 If μ is a measure, then for any continuous bounded function f

$$|\langle \mu | f \rangle| \leq C \parallel f \parallel_\infty . \tag{3.13}$$

We pick

$$f = \sum_k \varepsilon_{j,k} \psi_{j,k}, \qquad \text{where} \quad \varepsilon_{j,k} = \pm 1,$$

so that f is continuous and satisfies $\parallel f \parallel_\infty \leq C$, where C depends only on the wavelet (but not on the choice of the $\varepsilon_{j,k}$). Denoting by $c_{j,k}$ the wavelet coefficients of μ, we have

$$\langle \mu | f \rangle = \sum_k \varepsilon_{j,k} \int \psi_{j,k} d\mu = 2^{-j} \sum_k \varepsilon_{j,k} c_{j,k}.$$

Picking $\varepsilon_{j,k} = \text{sgn}(c_{j,k})$ it follows from (3.13) that

$$2^{-j} \sum_k |c_{j,k}| \leq C, \tag{3.14}$$

or, in other words, μ belongs to the Besov space $B_1^{0,\infty}$, which implies that $\eta_\mu(1) \geq 0$, see [14, 22].

On other hand, if $\mu \notin L^1$, then using the interpretation of the scaling function in terms of Sobolev spaces given by (3.2), we obtain that $\eta_\mu(1) \leq 0$. Hence the first part of the proposition holds.

We now prove (3.11). We assume that the used wavelet is compactly supported, and that its support is included in the interval $[-2^l, 2^l]$ for an $l > 0$ (we pick the smallest l such that this is possible). Let $\delta > \overline{\dim}_B(supp(\mu))$; for j large enough, $supp(\mu)$ is included in at most $2^{[\delta j]}$ intervals of length 2^{-j}. It follows that, at scale j, there exist at most $2^{[\delta j]} \cdot 2 \cdot 2^l$ wavelets $(\psi_{j,k})_{k \in \mathbb{Z}}$ whose support intersects the support of μ. Thus for j large enough, there are at most $C 2^{\delta j}$ wavelet coefficients that do not vanish.

Let $p \in (0, 1)$, $q = 1/p$ and r be the conjugate exponent of q, i.e. such that $1/q + 1/r = 1$. Using Hölder's inequality,

$$\sum_k |c_{j,k}|^p \leq \left(\sum_k |c_{j,k}|^{pq} \right)^{1/q} \left(\sum_k 1^r \right)^{1/r},$$

where the sums are over at most $C 2^{\delta j}$ terms; thus

$$\sum_k |c_{j,k}|^p \leq \left(\sum_k |c_{j,k}| \right)^p C\, 2^{\delta j / r}.$$

Using (3.14), we obtain that

$$2^{-j} \sum_k |c_{j,k}|^p \leq C 2^{-(1-\delta) j / r},$$

so that $\eta_\mu(p) \geq (1 - \delta)(1 - p)$. Since this is true $\forall \delta > \delta_\mu$, (3.11) follows.

We now prove (3.12). Let $p \geq 1$ and let q be the conjugate exponent. Using Hölder's inequality,

$$\sum_k |c_{j,k}| \leq \left(\sum_k |c_{j,k}|^p \right)^{1/p} \left(\sum_k 1^q \right)^{1/q}.$$

Let again $\delta > \delta_\mu$; using the fact that the sums bear on at most $2^{\delta j}$ terms, and that the left-hand side is larger than $C2^j$, we obtain that

$$\left(\sum_k |c_{j,k}|^p \right)^{1/p} \geq C\, 2^j 2^{-\delta j/q},$$

which can be rewritten

$$2^{-j} \sum_k |c_{j,k}|^p \geq C\, 2^{-j} 2^{pj} 2^{-\delta jp/q},$$

so that $\eta_\mu(p) \leq (1-p)(1-\delta)$; since this is true $\forall \delta > \delta_\mu$, (3.12) follows, and Proposition 3.6 is completely proved.

Since $p = 1$ is a borderline case for the use of the 1-exponent one may expect that picking $p < 1$ would yield $\eta_\mu(p) > 0$ (in which case one would be on the safe side in order to recover mathematical results concerning the p-spectrum, see [2, 15]). However, this is not the case, since there exist even continuous functions f that satisfy $\forall p > 0$, $\eta_f(p) = 0$. An example is supplied by

$$f = \sum_{j \geq 0} \sum_{k \in \mathbb{Z}} \frac{1}{j^2} \psi_{j,k}.$$

4 Multifractal Analysis of Lacunary Wavelet Series

Multifractal analysis is motivated by the observation that many mathematical models have an extremely erratic pointwise regularity exponent which jumps everywhere; this is the case e.g. of multiplicative cascades or of Lévy processes, whose exponents h satisfy that

$$\text{a.s.} \quad \forall x_0, \qquad \limsup_{x \to x_0} h(x) - \liminf_{x \to x_0} h(x) \tag{4.1}$$

is bounded from below by a fixed positive quantity (we will see that this is also the case for lacunary wavelet series). This clearly excludes the possibility of any robust direct estimations of h. The driving idea of multifractal analysis is that one should rather focus on alternative quantities that

- are numerically computable on real life data in a stable way,
- yield information on the erratic behavior of the pointwise exponent.

Furthermore, for standard random models (such as the ones mentioned above) we require these quantities not to be random (i.e. not to depend on the sample path which is observed) but to depend on the characteristic parameters of the model only. The relationship between the **multifractal spectrum** and scaling functions (initially pointed out by U. Frisch and G. Parisi in [23]; see (4.6) below) satisfies these requirements.

We now recall the notion of multifractal spectrum. We denote by $\dim(A)$ the Hausdorff dimension of the set A.

Definition 4.1 Let $h(x)$ denote a pointwise exponent. The multifractal spectrum $d(H)$ associated with this pointwise exponent is

$$d(H) = \dim\{x : \quad h(x) = H\}.$$

In the case of the p-exponent, the sets of points with a given p-exponent will be denoted by $F_f^p(H)$:

$$F_f^p(H) = \{x_0 : \ h_f^p(x_0) = H\}, \tag{4.2}$$

and the corresponding multifractal spectrum (referred to as the p-spectrum) is denoted by $d^p(H)$; in the case of the lacunarity exponent, we denote it by $d^{\mathcal{L}}(L)$.

4.1 Derivation of the Multifractal Formalism

We now recall how $d(H)$ is expected to be recovered from global quantities effectively computable on real-life signals (following the seminal work of G. Parisi and U. Frisch [23] and its wavelet leader reinterpetation [14]). A key assumption is that this exponent can be derived from nonnegative quantities (which we denote either by $e_{j,k}$ or e_λ), which are defined on the set of dyadic intervals, by a log-log plot regression:

$$h(x_0) = \liminf_{j \to +\infty} \frac{\log\left(e_{\lambda_j(x_0)}\right)}{\log(2^{-j})}. \tag{4.3}$$

It is for instance the case of the p-exponent, as stated in (3.4) or (3.9), for which the quantities e_λ are given by the p-leaders d_λ^p.

In the case of the lacunarity exponent, quantities e_λ can be derived as follows: Let $\Delta q > 0$ small enough be given. If f has a $1/q$-exponent H and a lacunarity exponent L at x_0 then its $1/q$-leaders satisfy

$$d_j^{1/q}(x_0) \sim 2^{-Hj},$$

and its $1/(q+\Delta q)$-leaders satisfy

$$d_j^{1/(q+\Delta q)}(x_0) \sim 2^{-(H+\Delta q L)j};$$

we can eliminate H from these two quantities by considering the **$\mathcal{L}$-leaders:**

$$d_\lambda^{\mathcal{L}} := \left(\frac{d_j^{1/(q+\Delta q)}}{d_j^{1/q}} \right)^{1/\Delta q} \sim 2^{-Lj}.$$

(this argument follows a similar one developed in [20, Ch. 4.3] for the derivation of a multifractal analysis associated with the oscillation exponent).

The multifractal spectrum will be derived from the following quantities, referred to as the *structure functions*, which are similar to the ones that come up in the characterization of the wavelet scaling function in (3.3):

$$S_f(r,j) = \left(2^{-j} \sum_k |e_{j,k}|^r \right).$$

The scaling function associated with the collection of (e_λ) is

$$\forall r \in \mathbb{R}, \qquad \zeta_f(r) = \liminf_{j \to +\infty} \frac{\log\left(S_f(r,j)\right)}{\log(2^{-j})}. \tag{4.4}$$

Let us now sketch the heuristic derivation of the multifractal formalism; (4.4) means that, for large j,

$$S_f(r,j) \sim 2^{-\zeta(r)j}.$$

Let us estimate the contribution to $S_f(r,j)$ of the dyadic intervals λ that cover the points of E_H. By definition of E_H, they satisfy $e_\lambda \sim 2^{-Hj}$; by definition of $d(H)$, since we use cubes of the same width 2^{-j} to cover E_H, we need about $2^{d(H)j}$ such cubes; therefore the corresponding contribution is of the order of magnitude of

$$2^{-j} 2^{d(H)j} 2^{-Hrj} = 2^{-(1-d(H)+Hr)j}.$$

When $j \to +\infty$, the dominant contribution comes from the smallest exponent, so that

$$\zeta(r) = \inf_H (1 - d(H) + Hr). \tag{4.5}$$

By construction, the scaling function $\zeta(r)$ is a concave function on $\mathbb{R}$, see [14, 23, 24] which is in agreement with the fact that the right-hand side of (4.5) necessarily is a concave function (as an infimum of a family of linear functions) no matter whether $d(H)$ is concave or not. If $d(H)$ also is a concave function, then the Legendre transform in (4.5) can be inverted (as a consequence of the duality of convex functions), which justifies the following assertion.

Definition 4.2 A nonnegative sequence (e_λ), defined on the dyadic intervals, follows the multifractal formalism if the associated multifractal spectrum $d(H)$ satisfies

$$d(H) = \inf_{r \in \mathbb{R}} (1 - \zeta(r) + Hr). \tag{4.6}$$

The derivation given above is not a mathematical proof, and the determination of the range of validity of (4.6) (and of its variants) is one of the main mathematical problems concerning multifractal analysis. If it does not hold in complete generality, the multifractal formalism nevertheless yields an upper bound of the spectrum of singularities, see [14, 23, 24]: As soon as (4.3) holds,

$$d(H) \leq \inf_{r \in \mathbb{R}} (1 - \zeta(r) + Hr).$$

In applications, multifractal analysis is often used only as a classification tool in order to discriminate between several types of signals; then, one is not directly concerned with the validity of (4.6) but only with a precise computation of the new *multifractal parameters* supplied by the scaling function, or equivalently its Legendre transform. Note that studies of multifractality for the p-exponent have been performed by A. Fraysse who proved genericity results of multifractality for functions in Besov or Sobolev spaces in [10].

4.2 Description of the Model and Global Regularity

In this section, we extend to possibly negative exponents the model of lacunary wavelet series introduced in [12]. We assume that ψ is a wavelet in the Schwartz class (see however the remark after Theorem 4.6, which gives sufficient conditions of validity of the results of this section when wavelets of limited regularity are used). Lacunary wavelet series depend on a **lacunarity parameter** $\eta \in (0, 1)$ and a **regularity parameter** $\alpha \in \mathbb{R}$. At each scale $j \geq 0$, the process $X_{\alpha,\eta}$ has exactly $[2^{\eta j}]$ nonvanishing wavelet coefficients on each interval $[l, l + 1)$ ($l \in \mathbb{Z}$), their common size is $2^{-\alpha j}$, and their locations are picked at random: In each interval $[l, l + 1)$ ($l \in \mathbb{Z}$), all drawings of $[2^{\eta j}]$ among the 2^j possibilities $\frac{k}{2^j} \in [l, l + 1)$ have the same probability. Such a series is called a **lacunary wavelet series** of parameters (α, η). Note that, since α can be arbitrarily negative, $X_{\alpha,\eta}$ can actually be a random distribution of arbitrary large order. By construction

$$H^{min}_{X_{\alpha,\eta}} = \alpha,$$

and, more precisely, the sample paths of $X_{\alpha,\eta}$ are locally bounded if and only if $\alpha > 0$. The case considered in [12] dealt with $\alpha > 0$, and was restricted to the computation of Hölder exponents. Considering p-exponents allows us to extend the model to negative values of α, and also to see how the global sparsity of the wavelet expansion (most wavelet coefficients vanish) is related with the pointwise lacunarity of the sample paths. Note that extensions of this model in different directions have been worked out in [5, 9].

Since we are interested in local properties of the process X, we restrict our analysis to the interval $[0, 1)$ (the results proved in the following clearly do not depend on the particular interval which is picked); we can therefore assume that $k \in \{0, \cdots 2^j - 1\}$.

We first determine how α and η are related with the global regularity of the sample paths. The characterization (3.3) implies that the wavelet scaling function is given by

$$\forall p > 0, \qquad \eta_{X_{\alpha,\eta}}(p) = \alpha p - \eta + 1. \tag{4.7}$$

It follows that

$$p_0 := p_0(X_{\alpha,\eta}) = \begin{cases} \frac{\eta-1}{\alpha} & \text{if } \alpha < 0 \\ +\infty & \text{if } \alpha > 0. \end{cases}$$

Note that p_0 always exists and is strictly positive, even if α takes arbitrarily large negative values. We recover the fact that p-exponents allow us to deal with singularities of arbitrarily large negative order. We will see that this is a particular occurrence of a general result, see Proposition 5.2; the key property here is the sparsity of the wavelet series.

4.3 Estimation of the p-Leaders of $X_{\alpha,\eta}$

An important step in the determination of the p-exponent of sample paths of $X_{\alpha,\eta}$ at every point is the estimation of their p-leaders. We now assume that $p < p_0$, so that the sample paths of $X_{\alpha,\eta}$ locally belong to L^p and the p-exponent of $X_{\alpha,\eta}$ is well-defined everywhere. Recall that the p-leaders are defined by

$$l_\lambda = \left(\sum_{\lambda' \subset 3\lambda} |c_{\lambda'}|^p 2^{-(j'-j)} \right)^{1/p}. \tag{4.8}$$

The derivation of the p-exponent of $X_{\alpha,\eta}$ everywhere will be deduced from the estimation of the size of the p-leaders of $X_{\alpha,\eta}$. A key result is supplied by the following proposition, which states that the size of the p-leaders of a lacunary wavelet series is correctly estimated by the size of the first nonvanishing wavelet coefficient of smaller scale that is met in the set $\{\lambda' : \quad \lambda' \subset 3\lambda\}$.

Proposition 4.3 *Let $\alpha \in \mathbb{R}$, $\eta \in (0, 1)$ and let $X_{\alpha,\eta}$ be a lacunary wavelet series of parameters (α, η); for each dyadic interval λ (of width 2^{-j}), we define j' ($= j'(\lambda)$) as the smallest random integer such that*

$$\exists \lambda' \subset 3\lambda \quad \textit{such that } |\lambda'| = 2^{-j'} \quad \textit{and} \quad c_{\lambda'} \neq 0.$$

Then, a.s. $\exists J$, $\exists C, C' > 0$ such that $\forall j \geq J$, $\forall \lambda$ of scale j

$$C 2^{-\alpha j'} 2^{-(j'-j)/p} \leq l_\lambda \leq C' 2^{-\alpha j'} 2^{-(j'-j)/p} j^{2/p}$$

Proof This result will be implied by the exponential decay rate $2^{-(j'-j)}$ that appears in the definition of p-leaders together with the lacunarity of the construction; we will show that exceptional situations where this would not be true (as a consequence of local accumulations of nonvanishing coefficients) have a small probability and ultimately will be excluded by a Borel-Cantelli type argument. We now make this argument precise. For that purpose, we will need to show that the sparsity of wavelet coefficients is uniform, which will be expressed by a uniform estimate on the maximal number of nonvanishing coefficients $c_{\lambda'}$ that can be found for λ' (at a given scale j') included in a given interval 3λ. Such an estimate can be derived by interpreting the choice of the nonvanishing wavelet coefficients in the construction of the model as a coarsening (on the dyadic grid) of an **empirical process**. Let us now recall this notion, and the standard estimate on the increments of the empirical process that we will need.

Let $N_j = [2^{\eta j}]$ denote the number of nonvanishing wavelet coefficients at scale j. We can consider that the corresponding dyadic intervals λ have been obtained first by picking at random N_j points in the interval $[0, 1]$ (these points are now N_j independent uniformly distributed random variables on $[0, 1]$), and then by associating to each point the unique dyadic interval of scale j to which it belongs. Let P_t^j be the process starting from 0 at $t = 0$, which is piecewise constant and which jumps by 1 at each random point thus determined. The family of processes

$$\alpha_t^j = \sqrt{N_j} \left(\frac{P_t^j}{N_j} - t \right) \tag{4.9}$$

is called an **empirical process** on $[0, 1]$. The size of the increments of the empirical process on a given interval yields information on the number of random points picked in this interval. If it is of length l, then the expected number of points is $l[2^{\eta j}]$, and the deviation from this average can be uniformly bounded using the following result of W. Stute which is a particular case of Lemma 2.4 of [26].

Lemma 4.4 *There exist two positive constants C_1' and C_2' such that, if* $0 < l < 1/8$, $N_j l \geq 1$ *and* $8 \leq A \leq C_1' \sqrt{N_j l}$,

$$\mathbb{P}\left(\sup_{|t-s|\leq l} |\alpha_t^j - \alpha_s^j| > A\sqrt{l} \right) \leq \frac{C_2'}{l} e^{-A^2/64}.$$

Rewritten in terms of P_t^j, this means that

$$\mathbb{P}\left(\sup_{|t-s|\leq l} |P_t^j - P_s^j - N_j(t-s)| > A\sqrt{N_j l} \right) \leq \frac{C_2'}{l} e^{-A^2/64}. \tag{4.10}$$

Recall that the assumption $\lambda' \subset 3\lambda$ implies that $3 \cdot 2^{-j} \geq 2^{-j'}$. We will apply Lemma 4.4 differently for small values of j' where the expected number of nonvanishing coefficients $c_{\lambda'}$ that can be found for λ' (at a given scale j') included in a given interval λ is very small, and the case of large j' where this number increases geometrically.

We first assume that

$$2^{-j'} \geq j^2 2^{-j/\eta}. \tag{4.11}$$

We pick intervals of length $l = j'^2 2^{-\eta j'}$ and, for the constant A in Stute's lemma, we pick $A = j$. Then (4.10) applied with $N = [2^{\eta j'}]$ yields that, with probability at least $1 - e^{-j^2}$, the number of intervals λ' of scale j' picked in such intervals is

$$2^{\eta j'} l + O(j^2) = O(j^2).$$

We now assume that

$$2^{-j'} \leq j^2 2^{-j/\eta}. \tag{4.12}$$

Then we pick intervals of length $l = 3 \cdot 2^{-j}$, and $A = j + j'$. Then (4.10) applied with $N = [2^{\eta j'}]$ yields that, with probability at least $1 - e^{-(j+j')^2}$, the number of intervals λ' of scale j' picked in such intervals is

$$2^{\eta j'} l + O((j+j')^2 \sqrt{2^{\eta j'} l}) \leq 2 \cdot 2^{\eta j'} l. \tag{4.13}$$

We are now ready to estimate the size of l_λ, assuming that all events described above happen (indeed, we note that the probabilities such that these events do not happen have a finite sum, so that, by the Borel-Cantelli lemma, they a.s. all occur for j large enough).

At scales j' which satisfy (4.11), if at least one of the $\lambda' \subset 3\lambda$ does not vanish, then there are at most j^2 of them, and the corresponding contribution to the sum in (4.8) lies between $|(c_{\lambda'})^p 2^{-(j-j')}|$ and $j^2|(c_{\lambda'})^p 2^{-(j-j')}|$. At scales j' which satisfy (4.12), the contribution of the wavelet coefficients of scale j' to the sum lies between $2^{\eta j' l}|(c_{\lambda'})^p 2^{-(j-j')}|$ and its double. Since $c_{\lambda'} = 2^{-\alpha j'}$, the condition $p < p_0$ implies that these quantities decay geometrically, so that the order of magnitude of the p-leader is given by the first non-vanishing term in the sum. Hence Proposition 4.3 holds.

4.4 p-Exponents and Lacunarity

We now derive the consequences of Proposition 4.3 for the determination of the p-exponents of $X_{\alpha,\eta}$ at every point. We first determine the range of p-exponents. First, note that all p-leaders have size at most $2^{-\alpha j}$, so that the p-exponent is everywhere larger than α. In the opposite direction, as a consequence of (4.13), every interval 3λ of scale j includes at least one nonvanishing wavelet coefficient at scale $j/\eta + (\log j)^2$; therefore, all p-leaders have size at least

$$2^{-\alpha\left(\frac{j}{\eta}+\log j\right)^2 - \frac{1}{p}\left(\frac{j}{\eta}-j+(\log j)^2\right)}.$$

It follows that the p-exponents are everywhere smaller than

$$H_{max} := \frac{\alpha}{\eta} + \left(\frac{1}{\eta} - 1\right)\frac{1}{p}. \tag{4.14}$$

We have thus obtained that

$$\text{a.s.} \qquad \forall p < p_0, \ \forall x_0 \in \mathbb{R}, \qquad \alpha \leq h^p_{X_{\alpha,\eta}}(x_0) \leq H_{max}.$$

For each j, let E^j_ω denote the subset of $[0, 1]$ composed of intervals 3λ ($\lambda \in \Lambda_j$) inside which the first nonvanishing wavelet coefficient is attained at a scale $l \leq [\omega j]$, and let

$$E_\omega = \limsup E^j_\omega.$$

Proposition 4.3 implies that, if $x_0 \notin E_\omega$, then, for j large enough, all wavelet leaders $l_{\lambda_j(x_0)}$ are bounded by

$$j^2 2^{-\alpha\frac{j}{\eta} - \frac{1}{p}\left(\frac{j}{\eta}-j\right)},$$

so that:

$$\text{if } x_0 \notin E_\omega, \text{ then} \qquad h^p_{X_{\alpha,\eta}}(x_0) \geq \alpha\omega + \frac{\omega - 1}{p}. \tag{4.15}$$

On other hand, if $x_0 \in E_\omega$, then there exists an infinite number of p-leaders $l_{\lambda_j(x_0)}$ larger than

$$2^{-\alpha\frac{j}{\eta} - \frac{1}{p}\left(\frac{j}{\eta} - j\right)},$$

so that:

$$\text{if } x_0 \in E_\omega, \text{ then} \qquad h^p_{X_{\alpha,\eta}}(x_0) \leq \alpha\omega + \frac{\omega - 1}{p}. \tag{4.16}$$

It follows from (4.15) and (4.16) that the sets of points where the p-exponent takes the value

$$H = \alpha\omega + \frac{\omega - 1}{p}$$

are the sets

$$H_\omega = \bigcap_{\omega' > \omega} E_{\omega'} - \bigcup_{\omega' < \omega} E_{\omega'}.$$

We have thus obtained the following result.

Proposition 4.5 *Let $\alpha \in \mathbb{R}$, $\eta \in (0, 1)$ and let $X_{\alpha,\eta}$ be a lacunary wavelet series of parameters (α, η). Let $p < p_0$; the sets of points with a given p-exponent are the sets*

$$F^p_{X_{\alpha,\eta}}(H) = H_\omega \quad \textit{for} \quad \omega = \frac{H + 1/p}{\alpha + 1/p};$$

and additionally, if $x_0 \in H_\omega$, then

$$\mathcal{L}_{X_{\alpha,\eta}}(x_0) = \omega - 1.$$

Remark We actually do not need the wavelet used to be in the Schwartz class for Theorem 4.6 to be true. One can verify that, if the uniform regularity of the wavelet is larger than $\max(|\alpha|, |H_{max}|)$, then all previous computations remain valid.

In order to determine the p-spectra and the lacunarity spectrum, one has to determine the Hausdorff dimensions of the sets H_ω. We note that these sets do not depend on α and on p, but only on the parameter ω and on the random drawing of the locations of the non-vanishing wavelet coefficients. When $\alpha > 0$, the dimensions of these sets (expressed in a slightly different way) were determined in [12], where it is shown that

$$\dim(H_\omega) = \eta\omega.$$

The following result follows.

Theorem 4.6 *Let $\alpha \in \mathbb{R}$, $\eta \in (0,1)$ and let $X_{\alpha,\eta}$ be a lacunary wavelet series of parameters (α,η); the p-spectrum of $X_{\alpha,\eta}$ is supported by the interval $[\alpha, H_{max}]$ and, on this interval,*

$$a.s. \quad \forall p < p_0, \quad \forall H, \qquad d^p(H) = \eta \frac{H + 1/p}{\alpha + 1/p}.$$

Furthermore, its lacunarity spectrum is given by

$$a.s. \quad \forall L \in [0, 1/\eta - 1], \qquad d^{\mathcal{L}}(L) = \eta(L+1).$$

Remark It is also shown in [12] that all the sets H_ω are everywhere dense, so that the quantity (4.1) is equal everywhere to $H_{max} - \alpha$.

For the sake of completeness, we now sketch how these dimensions can be computed. We start by estimating the size of E_ω. Note that the number of intervals 3λ which comprise E^j_ω is bounded by

$$[2^{\eta j}] + [2^{\eta(j+1)}] + \cdots + [2^{\eta[\omega j]}] \leq C 2^{\eta\omega j}.$$

Using these intervals for $j \geq J$ as an ε-covering, we obtain the following bound for the Hausdorff dimension of E_ω

$$\dim(E_\omega) \leq \eta\omega. \tag{4.17}$$

We now consider the sets H_ω; it follows from (4.15) and (4.16) that

$$H_\omega = \bigcap_{\omega' > \omega} E_{\omega'} - \bigcup_{\omega' < \omega} E_{\omega'}.$$

Since $\forall \omega' < \omega$, $H_\omega \subset E_{\omega'}$, it follows from (4.16) that

$$\dim(H_\omega) \leq \eta\omega.$$

In order to get a lower bound on the Hausdorff dimension of H_ω, we will need the following (slightly) modified notion of δ-dimensional Hausdorff measure.

Definition 4.7 Let $A \subset \mathbb{R}$. For $\varepsilon > 0$ and $\delta \in [0, 1]$, let

$$M_{\varepsilon}^{\delta,\gamma}(A) = \inf_R \left(\sum_i |A_i|^{\delta} |\log(|A_i|)|^{\gamma} \right),$$

where R denotes an ε-covering of A, and where the infimum is taken on all ε-coverings. The (δ, γ)-dimensional Hausdorff measure of A is

$$M^{\delta,\gamma}(A) = \lim_{\varepsilon \to 0} M_{\varepsilon,}^{\delta,\gamma}(A). \tag{4.18}$$

Since E_{ω}^{j} is composed of $\sim C2^{\eta\omega j}$ randomly located intervals of length $3 \cdot 2^{-j}$, standard ubiquity arguments (such as in [7, 12]) yield that

$$M^{\eta\omega,2}(G_{\omega}) > 0;$$

(4.16) implies that $\bigcup_{\omega' < \omega} E_{\omega'}$ (which can be rewritten as a countable union) has a vanishing $(\eta\omega, 2)$-dimensional Hausdorff measure. Thus

$$M^{\eta\omega,2}\left(E_{\omega} - \bigcup_{\omega' < \omega} E_{\omega'} \right) > 0.$$

Since this set is included in H_{ω}, we obtain that

$$\dim(H_{\omega}) \geq \eta\omega.$$

It suffices now to rewrite these dimensions as a function of the p-exponent to obtain Theorem 4.6.

Numerical examples for the estimation of $d^p(H)$ and $d^{\mathcal{L}}(H)$ of a lacunary wavelet series are given in Fig. 5. As predicted by theory, the numerical estimates of the p-exponent multifractal spectra are not invariant with p but follow the evolution with p of the theoretical spectra $d^p(H)$. The positions of the mode of the estimated spectra have a constant negative bias; yet, quantitatively, they very well reproduce the shift of the mode of the theoretical spectra to smaller values of H for increasing p, revealing the lacunary nature of the function. A refined analysis is possible with the estimated lacunarity exponent multifractal spectrum $d^{\mathcal{L}}(H)$, which has been computed here for several values of p for illustration purposes. The mode of the spectrum is estimated at $H \approx 0.2$ (instead of the theoretical $H = 0.25$). This clearly indicates the existence of positive lacunarity exponents. While the estimates for small values of p fall short of revealing the full support of the theoretical multifractal spectrum, they still enable one to identify a relatively large interval of positive lacunarity exponent values. The best estimate of $d^{\mathcal{L}}(H)$ is obtained

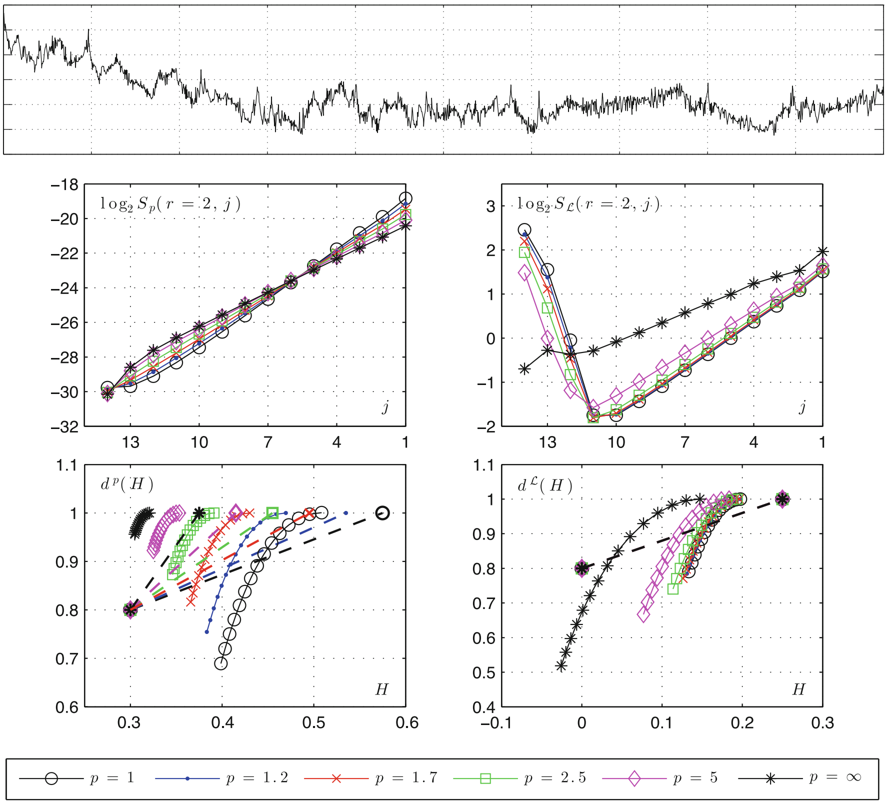


Fig. 5 **Lacunary wavelet series**: A typical sample path of a lacunary wavelet series ($\alpha = 0.3$, $\eta = 0.8$, *top row*) and estimated structure functions (*center row*) and multifractal spectra (*bottom row*) for p-exponents (*left column*) and lacunarity exponents (*right column*) obtained with different values of p. The *dashed lines* indicate the theoretical multifractal spectra

for the canonical value $p = p_0 = +\infty$ ($q = q_0 = 0$) in this example and produces a satisfactory concave envelope of the theoretical multifractal spectrum that provides clear evidence for ensembles of lacunary singularities with a range of positive exponents.

5 Concluding Remarks

The analysis that we developed is based on the assumption that $p_0(f) > 0$, or that $\eta_f(p) > 0$ for p small enough, so that p-exponents can be defined, at least, for $p \leq p_0$; we saw that this assumption allows us to deal with distributions of arbitrarily large order and, equivalently, to model pointwise singularities with arbitrarily large negative exponent. However, this does not imply that any tempered distribution satisfies these assumptions. Simple counterexamples are supplied by the **Gaussian fractional noises** B_α for $\alpha < 0$ whose sample paths can be seen as fractional derivatives of order $\frac{1}{2} - \alpha$ of the sample paths of a Brownian motion on $\mathbb{R}$ (Gaussian white noise corresponds to $\alpha = -1/2$, in which case it is a derivative, in the sense of distributions, of Brownian motion). In [18] the wavelet and leader scaling functions are derived, and it is proved that $\eta_{B_\alpha} = -\alpha p$, hence always is negative. However, the following result shows that, as soon as the wavelet expansion of the data has some sparsity, then this phenomenon no more occurs, and p_0 is always strictly positive (note that this situation is quite common in practice since sparse wavelet expansions are often met in applications).

Definition 5.1 A wavelet series $\sum_{j,k} c_{j,k}\psi_{j,k}$ is sparse if there exist $C > 0$ and $\eta < 1$ such that, on any interval $[l, l+1]$,

$$Card\{k : \; c_{j,k} \neq 0\} \leq C2^{\eta j}.$$

Typical examples of sparse wavelet series are supplied by lacunary wavelet series or by the measures which satisfy (3.10). The following proposition implies that multifractal analysis based on p-exponents is always possible for data with a sparse wavelet expansion.

Proposition 5.2 *Let f be a tempered distribution, which has a sparse wavelet expansion, then $\eta_f(p) > 0$ for p small enough, so that $p_0(f) > 0$.*

Proof Since f is a tempered distribution, it has a finite order, and thus it is a derivative of order A of a continuous function. Therefore f belongs to $C^{-A}(\mathbb{R})$, so that

$$|c_{j,k}| \leq C2^{Aj}.$$

Using again compactly supported wavelets, the same argument as in the proof of Proposition 3.6 yields that there are at most $C2^{\eta j}$ nonvanishing wavelet coefficients at scale j; it follows that

$$2^{-j} \sum_k |c_{j,k}|^p \leq C2^{-j}2^{\eta j}2^{Apj}$$

so that $\eta_f(p) \geq 1 - \eta - Ap$, and $\eta_f(p) > 0$ for $p < (1-\eta)/A$.

References

1. P. Abry, H. Wendt, S. Jaffard, H. Helgason, P. Goncalves, E. Pereira, C. Gharib, P. Gaucherand, M. Doret, Methodology for multifractal analysis of heart rate variability: from LF/HF ratio to wavelet leaders, in *Nonlinear Dynamic Analysis of Biomedical Signals EMBC conference (IEEE Engineering in Medicine and Biology Conferences)*, Buenos Aires (2010)

2. P. Abry, S. Jaffard, H. Wendt, A bridge between geometric measure theory and signal processing: multifractal analysis, in *Operator-Related Function Theory and Time-Frequency Analysis, The Abel Symposium 2012*, ed. by K. Grochenig, Y. Lyubarskii, K. Seip (Springer, Cham, 2015), pp. 1–56
3. P. Abry, S. Jaffard, H. Wendt, Irregularities and scaling in signal and image processing: multifractal analysis, in *Benoit Mandelbrot: A Life in Many Dimensions*, ed. by M. Frame, N. Cohen (World Scientific publishing, Singapore, 2015), pp. 31–116
4. A. Arneodo, E. Bacry, S. Jaffard, J.-F. Muzy, Singularity spectrum of multifractal functions involving oscillating singularities. J. Four. Anal. Appl. **4**, 159–174 (1998)
5. J.M. Aubry, S. Jaffard, Random wavelet series. Commun. Math. Phys. **227**(3), 483–514 (2002)
6. B. Barański, K. Bárány, J. Romanowska, On the dimension of the graph of the classical Weierstrass function. Adv. Math. **265**, 32–59 (2014)
7. V.V. Beresnevitch, S.S. Velani, A mass transference principle and the Duffin-Schaeffer conjecture for Hausdorff measures. Ann. Math. **164**, 971–992 (2006)
8. A.P. Calderon, A. Zygmund, Local properties of solutions of elliptic partial differential equations. Stud. Math. **20**, 171–223 (1961)
9. A. Durand, Random wavelet series based on a tree-indexed Markov chain. Commun. Math. Phys. **283**(2), 451–477 (2008)
10. A. Fraysse, Regularity criteria of almost every function in a Sobolev space. J. Funct. Anal. **258**, 1806–1821 (2010)
11. S. Jaffard, Multifractal formalism for functions. SIAM J. Math. Anal. **28**(4), 944–998 (1997)
12. S. Jaffard, On lacunary wavelet series. Ann. Appl. Probab. **10**(1), 313–329 (2000)
13. S. Jaffard, Pointwise regularity criteria. C. R., Math., Acad. Sci. Paris **339**(11), 757–762 (2004)
14. S. Jaffard, Wavelet techniques in multifractal analysis, in *Fractal Geometry and Applications: A Jubilee of Benoît Mandelbrot*, ed. by M. Lapidus, M. van Frankenhuijsen. Proceedings of Symposia in Pure Mathematics, vol. 72(2) (AMS, Providence, 2004), pp. 91–152
15. S. Jaffard, Pointwise regularity associated with function spaces and multifractal analysis, in *Approximation and Probability*, ed. by T. Figiel, A. Kamont. Banach Center Publications, vol. 72 (Institute of Mathematics, Polish Academy of Sciences, Warszawa, 2006), pp. 93–110
16. S. Jaffard, Wavelet techniques for pointwise regularity. Ann. Fac. Sci. Toulouse **15**(1), 3–33 (2006)
17. S. Jaffard, C. Melot, Wavelet analysis of fractal boundaries. Commun. Math. Phys. **258**(3), 513–565 (2005)
18. S. Jaffard, B. Lashermes, P. Abry, Wavelet leaders in multifractal analysis. in *Wavelet Analysis and Applications*, ed. by T. Qian, M.I. Vai, X. Yuesheng (Birkhäuser, Basel, 2006), pp. 219–264
19. S. Jaffard, P. Abry, S.G. Roux, B. Vedel, H. Wendt, The Contribution of Wavelets in Multifractal Analysis, in *Wavelet Methods in Mathematical Analysis and Engineering*, ed. by A. Damlamian, S. Jaffard. Series in Contemporary Applied Mathematics, vol. 14 (World Scientific publishing, Singapore, 2010), pp. 51–98
20. S. Jaffard, P. Abry, S.G. Roux, Function spaces vs. scaling functions: tools for image classification. in *Mathematical Image Processing*, ed. by M. Bergounioux. Springer Proceedings in Mathematics, vol. 5 (Springer, Berlin/Heidelberg/New York, 2011), pp. 1–39
21. S. Jaffard, C. Melot, R. Leonarduzzi, H. Wendt, S.G. Roux, M.E. Torres, P. Abry, p-Exponent and p-Leaders, Part I: negative pointwise regularity (2015). In review
22. Y. Meyer, *Ondelettes et Opérateurs*. (Hermann, Paris, 1990). English translation, *Wavelets and Operators*, (Cambridge University Press, 1992)
23. G. Parisi, U. Frisch, Fully developed turbulence and intermittency, in *Turbulence and Predictability in Geophysical Fluid Dynamics and Climate Dynamics*, ed. by M. Ghil, R. Benzi, G. Parisi. Proceedings of the International School (North-Holland, Amsterdam, 1985), p. 84
24. R.H. Riedi, Multifractal processes, in *Theory and Applications of Long Range Dependence*, ed. by P. Doukhan, G. Oppenheim, M.S. Taqqu (Birkhäuser, Boston, 2003), pp. 625–717
25. D. Rockmore, J. Coddington, J. Elton, Y. Wang, Multifractal analysis and authentication of Jackson Pollock paintings, Proc. SPIE 6810, 68100F-68100F-12 (2008)

26. W. Stute, The oscillation behavior of empirical processes. Ann. Probab. **10**, 86–107 (1982)
27. H. Triebel, Fractal characteristics of measures; an approach via function spaces. J. Four. Anal. Appl. **9**(4), 411–430 (2003)
28. H. Wendt, S.G. Roux, P. Abry, S. Jaffard, Wavelet leaders and bootstrap for multifractal analysis of images. Signal Process. **89**, 1100–1114 (2009)

Part V
Random Constructions

Dimensions of Random Covering Sets

Esa Järvenpää and Maarit Järvenpää

To our teacher Pertti Mattila, the inspiring source of our interest in geometric measure theory.

Abstract In this overview we discuss recent results on dimensional properties of random covering sets.

Keywords Random covering set • Limsup set • Hausdorff and packing dimensions

Mathematics Subject Classification (2000). Primary 60D05; Secondary 28A20

1 Introduction

1.1 Limsup Sets

Various types of limsup sets defined in a natural manner as upper limits of sequences of sets play an important role in many fields of mathematics. One of the well-known examples of such sets is related to the Diophantine approximation. According to the Dirichlet's approximation theorem, for all real numbers $x \in \mathbb{R}$, there are infinitely many positive integers $q \in \mathbb{N}$ satisfying $|x - \frac{p(q)}{q}| < \frac{1}{q^2}$ for some $p(q) \in \mathbb{Z}$. Replacing $\frac{1}{q^2}$ with a general function $\phi(q)$, leads to the following type of extension which has been extensively investigated in the literature:

$$|x - \frac{p(q)}{q}| < \phi(q). \tag{1.1}$$

E. Järvenpää • M. Järvenpää (✉)
Department of Mathematical Sciences, University of Oulu, P.O. Box 3000 90014, Oulu, Finland
e-mail: esa.jarvenpaa@oulu.fi; maarit.jarvenpaa@oulu.fi

C. Bandt et al. (eds.), *Fractal Geometry and Stochastics V*, Progress in Probability 70, DOI 10.1007/978-3-319-18660-3_16

The natural questions addressed in this context are as follows: Which properties of ϕ guarantee that for all or for almost all $x \in \mathbb{R}$ there are infinitely many $q \in \mathbb{N}$ satisfying (1.1) for some $p(q) \in \mathbb{Z}$? Given ϕ, what is the size of the set of points $x \in \mathbb{R}$ for which there exist infinitely many $q \in \mathbb{N}$ satisfying (1.1) for some $p(q) \in \mathbb{Z}$? Alternatively, one may study simultaneous approximations for $x_1, \ldots, x_n \in \mathbb{R}$. We refer to the work of Beresnevich and Velani [2] for a breakthrough in this direction.

Multiplying (1.1) by q gives the inequality

$$|xq - p(q)| < \tilde{\phi}(q) \tag{1.2}$$

which can be interpreted by means of rotations. Indeed, denoting by $R_x : \mathbb{T}^1 \to \mathbb{T}^1$ the rotation by the angle x on the circle $\mathbb{T}^1 := \mathbb{R}/\mathbb{Z}$, inequality (1.2) is satisfied if and only if the distance between zero and the q-th iterate of zero $R_x^q(0)$ is less than $\tilde{\phi}(q)$. Hence, the above mentioned problems are equivalent to the investigation of the size of the set

$$\{x \in \mathbb{R} \mid |R_x^q(0)| < \tilde{\phi}(q) \text{ for infinitely many } q \in \mathbb{N}\}$$

which can be regarded as a special case of the shrinking target problem or the dynamical Diophantine approximation formulated in full generality as follows: given a dynamical system $T : X \to X$ on a metric space X, a point $x_0 \in X$ and a sequence (r_n) of positive real numbers tending to zero, determine the size of the set

$$\{x \in X \mid T^n(x) \in B(x_0, r_n) \text{ for infinitely many } n \in \mathbb{N}\},$$

where $B(x, r)$ is the open ball with radius r centred at $x \in X$. A variant of this question is the moving target problem concerning the study of the following limsup set

$$\begin{aligned}\limsup_{n\to\infty} B(T^n(x_0), r_n) &= \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} B(T^n(x_0), r_n)\\ &= \{x \in X \mid x \in B(T^n(x_0), r_n) \text{ for infinitely many } n \in \mathbb{N}\}.\end{aligned}$$

We refer to [12] for a recent account on this line of research.

An interesting relation between limsup sets and Brownian motion was discovered by Khoshnevisan, Peres and Xiao in [23]. They defined a class of limsup random fractals for the purpose of characterising the sets which intersect the set of fast points almost surely. The notion of fast points was introduced by Orey and Taylor [28] as the set of times where the increments of the Brownian motion are exceptionally large.

1.2 Random Covering Sets

Random covering sets are a class of limsup sets defined by means of a family of randomly distributed subsets of the d-dimensional torus $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$. More precisely, letting $(A_n)_{n=1}^{\infty}$ be a sequence of non-empty subsets of $\mathbb{T}^d$ and letting $\mathbf{x} = (x_n)_{n=1}^{\infty}$ be a sequence of independent random variables which are uniformly distributed on $\mathbb{T}^d$, define the random covering set $E = E(\mathbf{x})$ by

$$E(\mathbf{x}) = \limsup_{n\to\infty}(x_n + A_n) = \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}(x_k + A_k),$$

where $x + A := \{x + y \mid y \in A\}$. In the case $A_k = B(0, r_k)$ the study of random covering sets may be interpreted as a moving target problem for the random walk on $\mathbb{T}^d$.

Denoting the Lebesgue measure on $\mathbb{T}^d$ by $\mathcal{L}$, it follows from the Borel-Cantelli lemma and Fubini's theorem that, almost surely, either $\mathcal{L}(E) = 0$ or $\mathcal{L}(E) = 1$ depending on whether the series $\sum_{k=1}^{\infty}\mathcal{L}(A_k)$ converges or diverges, respectively, that is, almost all or almost no points of the torus are covered, depending on whether or not the series of the Lebesgue measures of the generating sets diverges.

The case of full Lebesgue measure has been extensively studied. Even in the simplest case when $d = 1$ and the generating sets are intervals of length l_n it was a long-standing problem to find conditions which guarantee that the whole circle $\mathbb{T}^1$ is covered almost surely. This problem, known in literature as the Dvoretzky covering problem, was first posed by Dvoretzky [6] in 1956. After substantial contributions of many authors, including Billard [3], Erdős [8], Kahane [19] and Mandelbrot [26], the full answer to the Dvoretzky covering problem was given in this context by Shepp [30] in 1972. He verified that $E = \mathbb{T}^1$ almost surely if and only if

$$\sum_{n=1}^{\infty}\frac{1}{n^2}\exp(l_1 + \cdots + l_n) = \infty,$$

where the lengths (l_n) are in decreasing order.

In the higher dimensional case the Dvoretzky covering problem has been studied by El Hélou [7] and Kahane [21] among others. In [21] Kahane gave a complete solution for the problem when the generating sets are similar simplexes. However, in the general case the covering problem is still unsolved.

For an overview on various results concerning random covering sets and related topics, we refer to [20, Chapter 11] and [22] and the references therein. We briefly mention a few modifications to the random covering model. As a generalisation of the Dvoretzky covering problem one can consider the number of covering times for a given set $K \subset \mathbb{T}^d$, or whether $K \subset E$ almost surely. For different approaches to related questions, see [1, 7, 9, 10, 14, 16, 20] and [21]. The Dvoretzky covering problem can be formulated in terms of Mandelbrot cutout sets [27]. For the solution of the covering problem in this setting, see Shepp [31]. In metric spaces the random coverings by balls have been studied in [15]. Recent contributions to the topic include various types of dynamical models, see [12, 18] and [25], and projections of random covering sets [4].

2 Dimension Results in the Torus

From now on we focus on the natural problem of determining the almost sure value of the Hausdorff dimension of the covering set E in the case of zero Lebesgue measure. The investigation of dimensional properties of covering sets was pioneered by Fan and Wu [11]. They gave a formula for the almost sure Hausdorff dimension, denoted by $\dim_{\mathrm{H}}$, of the limsup set in the circle $\mathbb{T}^1$ provided that each generating set A_n is an open interval of length $l_n = 1/n^\alpha$ for $\alpha > 1$. Using different methods, Durand [5] studied the case of arbitrary decreasing sequences of lengths $(l_n)_{n=1}^\infty$ and proved the following generalisation of the dimension formula: almost surely

$$\dim_{\mathrm{H}} E = \inf\{t \geq 0 \mid \sum_{n=1}^{\infty} l_n^t < \infty\} = \limsup_{n\to\infty} \frac{\log n}{-\log l_n}. \tag{2.1}$$

Note that here the covering set E is almost surely a dense G_δ-set in $\mathbb{T}^1$ since $A_n =]0, l_n[$ is an open interval. This implies that both packing and Minkowski dimensions of E are equal to 1. When considering hitting probabilities of random covering sets in the circle, Li, Shieh and Xiao [24] gave an alternative proof of the dimension result (2.1) under additional assumptions.

The various methods used in the proofs of the above results rely heavily on the facts that the ambient space is 1-dimensional or the generating sets are ball like. The natural question of calculating the almost sure dimension value of the covering set in the d-dimensional torus $\mathbb{T}^d$ was first addressed in [17] for generating sets of the form $A_n = \Pi(L_n(R))$, where $\Pi : \mathbb{R}^d \to \mathbb{T}^d$ is the natural covering map, R is a subset of the closed unit cube $[0, 1]^d$ with non-empty interior and the map $L_n : \mathbb{R}^d \to \mathbb{R}^d$ is a contractive linear injection such that for all $i = 1, \ldots, d$ the sequence of singular values $(\sigma_i(L_n))_{n=1}^\infty$ decreases to 0 as n tends to infinity. Recall that $\sigma_i(L_n)$ is the length of i-th longest semiaxis of $L_n(B(0, 1))$. It turns out that, almost surely, the Hausdorff dimension of E can be calculated in terms of the singular value functions

$$\Phi^t(L_n) := \sigma_1(L_n) \cdots \sigma_{m-1}(L_n)\sigma_m(L_n)^{t-m+1}$$

where m is an integer with $m-1 < t \leq m$ and $0 < \sigma_d(L_n) \leq \cdots \leq \sigma_1(L_n) < 1$. More precisely, almost surely

$$\dim_{\mathrm{H}} E = \inf\{0 < t \leq d \mid \sum_{n=1}^{\infty} \Phi^t(L_n) < \infty\} \tag{2.2}$$

with the interpretation $\inf \emptyset = d$, see [17].

Recently, Persson [29] showed that for open generating sets $A_n \subset \mathbb{T}^d$, almost surely,

$$\dim_{\mathrm{H}} E \geq \inf\{0 < t \leq d \mid \sum_{n=1}^{\infty} g_t(A_n) < \infty\}, \tag{2.3}$$

where

$$g_t(A) = \frac{\mathcal{L}(A)^2}{I_t(A)}$$

for all Lebesgue measurable sets $A \subset \mathbb{T}^d$. Here

$$I_t(A) = \iint_{A \times A} |x-y|^{-t} \, d\mathcal{L}(x) \, d\mathcal{L}(y) \tag{2.4}$$

is the t-energy of A. Note that, for simplicity, we use the notation $|x-y|$ for both the Euclidean distance and the natural distance in $\mathbb{T}^d$.

It is straightforward to see that inequality (2.3) gives a generalisation of (2.2). Indeed, it follows easily that the lower bound in (2.3) equals the right-hand side of (2.2) under the assumption that the generating sets A_n are open rectangles. The fact that in [17] the monotonicity assumption $\sigma_i(L_n) \downarrow 0$ as $n \to \infty$ is not needed for the purpose of verifying the upper bound of the Hausdorff dimension in (2.2) implies that (2.2) is valid for generating sets of the form $A_n = \Pi(L_n(R))$ without the convergence assumption. To conclude, the results of this section cover the case of $A_n \subset \mathbb{T}^d$ being box-like, that is, linear images of a set having non-empty interior, and moreover, give a lower bound for general open sets.

3 Dimension Results in Riemann Manifolds

In this section we concentrate on dimensional properties of random covering sets in a d-dimensional Riemann manifold. The setting we are dealing with is quite general: the generating sets are Lebesgue measurable, and moreover, instead of the uniform distribution $\mathcal{L}$ we consider any non-singular measure. This section is based on our recent joint work with De-Jun Feng and Ville Suomala [13].

We begin by introducing the notation. Let $U, V \subset \mathbb{R}^d$ be open, simply connected and bounded sets, and let $\Xi : U \times V \to \mathbb{R}^d$ be a C^1-map such that the maps $\Xi(\cdot, y) : U \to \Xi(U, y)$ and $\Xi(x, \cdot) : V \to \Xi(x, V)$ are diffeomorphisms for all $(x, y) \in U \times V$. Denoting the derivatives of $\Xi(\cdot, y)$ and $\Xi(x, \cdot)$ by $D_1\Xi$ and $D_2\Xi$, respectively, we assume that there are constants $C_l, C_u > 0$ such that the singular values satisfy

$$C_l \leq \sigma_i(D_1\Xi(x, y)), \sigma_i(D_2\Xi(x, y)) \leq C_u \tag{3.1}$$

for all $(x, y) \in U \times V$ and for all $i = 1, \ldots, d$.

Let Γ be a Radon probability measure on U which is not purely singular with respect to the Lebesgue measure $\mathcal{L}$. We consider the probability space $(U^{\mathbb{N}}, \mathcal{F}, \mathbb{P})$ which is the completion of the infinite product of $(U, \mathcal{B}, \Gamma)$, where $\mathcal{B}$ is the Borel σ-algebra on U. Assuming that $A_n \subset V$ for all $n \in \mathbb{N}$, define the random covering set $E(\mathbf{x})$ for all $\mathbf{x} = (x_n)_{n \in \mathbb{N}} \in U^{\mathbb{N}}$ by

$$E(\mathbf{x}) := \limsup_{n \to \infty} \Xi(x_n, A_n) = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} \Xi(x_k, A_k).$$

Note that, choosing $U = V = \mathbb{T}^d$ and $\Xi(x, y) = x + y$, gives the setting of Sect. 2.

For the purpose of computing the almost sure value of the Hausdorff dimension of covering sets, we need the following quantities. For $0 \leq t < \infty$, the t-dimensional Hausdorff content of a set $A \subset \mathbb{R}^d$ is denoted by

$$\mathcal{H}^t_\infty(A) = \inf\{\sum_{n=1}^{\infty} (\operatorname{diam} B_n)^t \mid A \subset \bigcup_{n=1}^{\infty} B_n\},$$

where $\operatorname{diam} B$ is the diameter of a set $B \subset \mathbb{R}^d$. Define

$$t_0 := \inf\{0 \leq t \leq d \mid \sum_{n=1}^{\infty} \mathcal{H}^t_\infty(A_n) < \infty\} \tag{3.2}$$

with the interpretation $\inf \emptyset = d$. For Lebesgue measurable sets $A_n \subset \mathbb{R}^d$, we set

$$s_0 := \sup\{0 \leq s \leq d \mid \sum_{n=1}^{\infty} G_s(A_n) = \infty\}, \tag{3.3}$$

where

$$G_s(B) := \sup\{g_s(A) \mid A \subset B,\ A \text{ is Lebesgue measurable and } \mathcal{L}(A) > 0\}$$

with the interpretation $\sup \emptyset = 0$. Finally, we say that a point $x \in A \subset \mathbb{R}^d$ has *positive density* if

$$\liminf_{r\to 0} \frac{\mathcal{L}(A \cap B(x,r))}{\mathcal{L}(B(x,r))} > 0.$$

Now we are ready to state our main theorem from [13].

Theorem 3.1 *Suppose that $K \subset V$ is compact and the generating sets $(A_n)_{n=1}^{\infty}$ are subsets of K. Then*

(a) $\dim_{\mathrm{H}} E(\mathbf{x}) \le t_0$ *for all* $\mathbf{x} \in U^{\mathbb{N}}$,
(b) $\dim_{\mathrm{H}} E(\mathbf{x}) \ge s_0$ *for $\mathbb{P}$-almost all $\mathbf{x} \in U^{\mathbb{N}}$ provided that A_n is Lebesgue measurable for all $n \in \mathbb{N}$,*
(c) $\dim_{\mathrm{H}} E(\mathbf{x}) = s_0 = t_0$ *for $\mathbb{P}$-almost all $\mathbf{x} \in U^{\mathbb{N}}$ provided that for all $n \in \mathbb{N}$ the set A_n is analytic and all points in A_n have positive density,*
(d) $\dim_{\mathrm{p}} E(\mathbf{x}) = d$ *for $\mathbb{P}$-almost all $\mathbf{x} \in U^{\mathbb{N}}$ provided that A_n is Lebesgue measurable and $\mathcal{L}(A_n) > 0$ for all $n \in \mathbb{N}$.*

As a consequence of Theorem 3.1 we obtain the following dimension result for random covering sets in compact Riemann manifolds [13]. Note that in Corollary 3.2 the quantities t_0 and s_0 are defined similarly as in (3.2) and (3.3) by using the distance function induced by the Riemann metric and by replacing $\mathcal{L}$ by the Riemann volume.

Corollary 3.2 *Let M_1, M_2 and N be d-dimensional compact Riemann manifolds. Suppose that $\Xi : M_1 \times M_2 \to N$ is a C^1-map such that $\Xi(x,\cdot)$ and $\Xi(\cdot,y)$ are local diffeomorphisms satisfying (3.1). Let $(A_n)_{n=1}^{\infty}$ be a sequence of analytic subsets of M_2 such that all points in every A_n have positive density with respect to the Riemann volume on M_2. Assume that Γ is a Radon measure on M_1 which is not purely singular with respect to the Riemann volume on M_1. Then for $\mathbb{P}$-almost every $\mathbf{x} \in (M_1)^{\mathbb{N}}$, we have* $\dim_{\mathrm{H}} E(x) = t_0 = s_0$ *and* $\dim_{\mathrm{p}} E(\mathbf{x}) = d$.

In Corollary 3.2 the compactness assumption guarantees that the random covering set is non-empty. The claim is valid for compact subsets of non-compact Riemann manifolds as well. Note that Corollary 3.2 is sharp in the sense that, almost surely, $\dim_H E(\mathbf{x})$ may be strictly between s_0 and t_0 if we only assume that for all $n \in \mathbb{N}$ we have $\mathcal{L}(A_n \cap B(x,r)) > 0$ for all $x \in A_n$ and $r > 0$. Moreover, even for open generating sets the quantity s_0 may be strictly larger than the right hand side of (2.3). In particular, the lower bound in (2.3) is not the best possible one. Finally, Theorem 3.1 may fail if the distribution Γ is singular with respect to the Lebesgue measure. We refer to [13] for details of these facts.

Acknowledgements We thank the Centre of Excellence in Analysis and Dynamics Research funded by the Academy of Finland.

References

1. J. Barral, A.H. Fan, Covering numbers of different points in Dvoretzky covering. Bull. Sci. Math. **129**(4), 275–317 (2005)
2. V. Beresnevich, S. Velani, A mass transference principle and the Duffin-Schaeffer conjecture for Hausdorff measures. Ann. Math. (2) **164**(3), 971–992 (2006)
3. P. Billard, Séries de Fourier aléatoirement bornées, continues, uniformément convergentes. Ann. Sci. École Norm. Sup. (3) **82**, 131–179 (1965)
4. C. Chen, H. Koivusalo, B. Li, V. Suomala, Projections of random covering sets. J. Fractal Geom. **1**(4), 449–467 (2014)
5. A. Durand, On randomly placed arcs on the circle, in *Recent Developments in Fractals and Related Fields*. Applied and Numerical Harmonic Analysis (Birkhäuser, Boston, 2010), pp. 343–351
6. A. Dvoretzky, On covering a circle by randomly placed arcs. Proc. Nat. Acad. Sci. U.S.A. **42**, 199–203 (1956)
7. Y. El Hélou, Recouvrement du tore T^q par des ouverts aléatoires et dimension de Hausdorff de l'ensemble non recouvert. C. R. Acad. Sci. Paris Sér. A-B **287**, A815–A818 (1978)
8. P. Erdős, Some unsolved problems. Magyar Tud. Akad. Mat. Kutató Int. Közl. **6**, 221–254 (1961)
9. A.H. Fan, How many intervals cover a point in Dvoretzky covering? Isr. J. Math. **131**, 157–184 (2002)
10. A.-H. Fan, J.-P. Kahane, Rareté des intervalles recouvrant un point dans un recouvrement aléatoire. Ann. Inst. Henri Poincaré Probab. Stat. **29**, 453–466 (1993)
11. A.-H. Fan, J. Wu, On the covering by small random intervals. Ann. Inst. Henri Poincaré Probab. Stat. **40**, 125–131 (2004)
12. A.-H. Fan, J. Schmeling, S. Troubetzkoy, A multifractal mass transference principle for Gibbs measures with applications to dynamical Diophantine approximation. Proc. Lond. Math. Soc. **107**, 1173–1219 (2013)
13. D.-J. Feng, E. Järvenpää, M. Järvenpää, V. Suomala, Dimensions of random covering sets in Riemann manifolds (Manuscript)
14. J. Hawkes, On the covering of small sets by random intervals. Quart. J. Math. Oxf. Ser. (2) **24**, 427–432 (1973)
15. J. Hoffmann-Jørgensen, Coverings of metric spaces with randomly placed balls. Math. Scand. **32**, 169–186 (1973)
16. S. Janson, Random coverings in several dimensions. Acta Math. **156**, 83–118 (1986)
17. E. Järvenpää, M. Järvenpää, H. Koivusalo, B. Li, V. Suomala, Hausdorff dimension of affine random covering sets in torus. Ann. Inst. Henri Poincaré Probab. Stat. **50**, 1371–1384 (2014)
18. J. Jonasson, J.E. Steif, Dynamical models for circle covering: Brownian motion and Poisson updating. Ann. Probab. **36**, 739–764 (2008)
19. J.-P. Kahane, Sur le recouvrement d'un cercle par des arcs disposés au hasard. C. R. Acad. Sci. Paris **248**, 184–186 (1956)
20. J.-P. Kahane, *Some Random Series of Functions*. Cambridge Studies in Advanced Mathematics, vol. 5 (Cambridge University Press, Cambridge/New York, 1985)
21. J.-P. Kahane, *Recouvrements aléatoires et théorie du potentiel*. Colloq. Math. **60/61**, 387–411 (1990)
22. J.-P. Kahane, Random coverings and multiplicative processes, in *Fractal Geometry and Stochastics II*. Progress in Probability, vol. 46 (Birkhäuser, Basel/Boston, 2000), pp. 125–146
23. D. Khoshnevisan, Y. Peres, Y. Xiao, Limsup random fractals. Electron. J. Probab. **5**(5), 24pp (2000)
24. B. Li, N.-R. Shieh, Y. Xiao, Hitting probabilities of the random covering sets, in *Fractal Geometry and Dynamical Systems in Pure and Applied Mathematics. II. Fractals in Applied Mathematics*. Contemporary Mathematics, vol. 601 (American Mathematical Society, Providence, 2013), pp. 307–323

25. L. Liao, S. Seuret, Diophantine approximation by orbits of expanding Markov maps. Ergod. Theory Dyn. Syst. **33**, 585–608 (2012)
26. B. Mandelbrot, On Dvoretzky coverings for the circle. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **22**, 158–160 (1972)
27. B. Mandelbrot, Renewal sets and random cutouts. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **22**, 145–157 (1972)
28. S. Orey, S.J. Taylor, How often on a Brownian path does the law of iterated logarithm fail? Proc. Lond. Math. Soc. **28**, 174–192 (1974)
29. T. Persson, A note on random coverings of tori. Bull. Lond. Math. Soc. **47**(1), 7–12 (2015)
30. L.A. Shepp, Covering the circle with random arcs. Isr. J. Math. **11**, 328–345 (1972)
31. L.A. Shepp, Covering the line with random intervals. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **23**, 163–170 (1972)

Expected Lifetime and Capacity

András Telcs and Marianna E.-Nagy

Abstract We investigate sharp isoperimetric problems for random walks on weighted graphs. Symmetric weights on edges determine the one step transition probabilities for the random walk, measure of sets and capacity between sets. In that setup one can be interested in the exit time of the random walk from a set, i.e. to find for a fixed starting point the "optimal" set of given volume which maximizes the expected time when the walk leaves the set. A strongly related problem is to find a set of fixed volume which has minimal conductance with respect to a given set. In both problems the answer is less appealing than in the case of Euclidean space. As demonstrated by a simple counterexample, there is no unique optimal set. The Berman-Konsowa principle is used in the search for optimal sets. It allows to construct a new graph on which the calculation of conductance and mean exit time is tractable.

Keywords Isoperimetric inequality • Random walks • Berman-Konsowa principle

1 Introduction

Isoperimetric problems have a long and shining history in mathematics as well as in human culture. Pappus credited to Zenodorus the first statement of the two dimensional isoperimetric problem. Several other isoperimetric problems were formulated in the course of time. One can find a classical introduction in Polya's and Szegő's book [9] and further references in the nice survey of Caroll [5]. We are

A. Telcs (✉)
Department of Computer Science and Information Theory, Budapest University of Technology and Economics, Magyar tudósok körútja 2, 1117 Budapest, Hungary

Department of Quantitative Methods, University Pannonia, 8200 Veszprém, Hungary
e-mail: telcs.szit.bme@gmail.com

M. E.-Nagy
Department of Differential Equations, Budapest University of Technology and Economics, Egry József utca 1, 1111 Budapest, Hungary

C. Bandt et al. (eds.), *Fractal Geometry and Stochastics V*, Progress in Probability 70, DOI 10.1007/978-3-319-18660-3_17

not going to present a review here. To find the optimal set for the maximal expected lifetime of a planar Brownian motion in a finite closed, connected domain of fixed area is a naturally arising similar problem (cf. [1, 5] and their references). In the light of recent developments in the study of diffusion processes in measure metric Dirichlet spaces (cf. [3, 7]) it is natural to rise the same question on such spaces.

Let us imagine that we have a sheep, a piece of grassland, and an electric fence of a given length. The sheep starts at a given point at time zero and performs a diffusion according to a fixed measure and a local, regular, Markovian Dirichlet form. We want to enclose the sheep with the fence in such a way that the sheep is hit by electricity as late as possible, in expectation. This scenario inspired Erin Pearse to coin the name "Brownian sheep" at the Cornell Conference on Analysis and Probability on Fractals in 2005.

In the present paper we make a very first step towards the solution of the Brownian sheep problem. We consider a discrete space-time counterpart of the problem, given by random walks on weighted graphs (for general introduction and background c.f. [6, 10]).

We provide a characterization of the optimal solution for:

1. The minimal capacity problem: given two sets Γ and $F \subset \Gamma$ and a constant M, find a set $D \subset F$ with volume not larger than M such that the capacity between D and $\Gamma \backslash F$ is minimal.
2. The maximal lifetime problem: given a starting point of the walk and a constant M, find a set F of volume not larger than M such that the expected exit time of the walk is maximal.

The key tool for us is the Berman-Konsowa (B-K in the sequel) principle [4] (see also [8] for a nice interpretation), by which the problem can be reduced to star graphs.

The paper is organized as follows. In Sect. 2 we introduce basic notation and facts. In Sect. 3 we present the Berman-Konsowa principle, in Sect. 4 we discuss the problem of capacity and in Sect. 5 the problem of Brownian Sheep. Some technical details are collected in an Appendix.

2 Foundations

Let (Γ, Q_Γ, μ) be a connected weighted undirected graph with vertex set Γ, edge set Q_Γ and a symmetric weight on edges $\mu_{x,y} = \mu_{y,x}$. The corresponding resistance is $R_{x,y} = 1/\mu_{x,y}$. For $x \in \Gamma$ let $\mu(x) = \sum_y \mu_{x,y}$.

For sake of simplicity we will solve the problem on the *cable system* of the graph, i.e., all edges are considered as copies of the unit interval [2]. For an edge (x, y) and $\alpha \in [0, 1]$ let (α, x, y) denote the point which splits the edge into $\alpha, 1-\alpha$ parts. We write $w_0 = (0, x, y)$ for x, $w_1 = (1, x, y)$ for y, and $w_\alpha = (\alpha, x, y)$ for points on the edge. Resistance and weight are proportional to the length of a subinterval:

$$R_{w_0,w_\alpha} = \alpha R_{x,y} \text{ and } \quad \mu_{w_0,w_\alpha} = \alpha \mu_{x,y}.$$

The basic space for our study is the set of all points of the unit intervals representing edges. It is denoted by W. We consider subsets $A \subset W$ which are unions of subintervals where adjacent endpoints are identified. We assume that such a set A is convex in the following sense: if $w_{\bar{\alpha}} = (\bar{\alpha}, x, y) \in A$ then at least one of the vertices x and y is in A, and if lets say $x \in A$, then $w_\alpha = (\alpha, x, y) \in A$ for all $\alpha \in [0, \bar{\alpha}]$ as well.

In the sequel the investigated sets $A \subset W$ are assumed to be open and precompact. Let $\overline{A}$ denote the closure of the set and $\partial A = \overline{A} \backslash A$ the boundary of A. The boundary of a set is a discrete set of points on intervals. The set of edges crossing ∂A will be denoted by cA.

The weights on edges define a measure $d\mu(\alpha, x, y) = \mu_{x,y} d\alpha$, with

$$\mu(A) = \sum_{x,y \in A \cap \Gamma} \mu_{x,y} + \sum_{\substack{(\alpha,x,y) \in \partial A \\ x \in A, y \notin A}} \alpha \mu_{x,y}.$$

We consider the usual random walk $X_n \in \Gamma$ on (Γ, μ) defined by the transition probability $P(x, y) = \mu_{x,y} / \mu(x)$. We assume that there is a $p_0 > 0$ such that for all $(x, y) \in Q_\Gamma$

$$P(x, y) \geq p_0. \tag{2.1}$$

As a consequence $\deg(x) \leq 1/p_0$ for all $x \in \Gamma$, i.e. the graph has bounded degree. Γ can be infinite, however.

Now we define the killed random walk for a set A which contains a finite number of vertices. We assign to A a corresponding graph with vertex set $\Gamma_A = \Gamma \cap A \cup \partial A$ and the induced edges. On this graph we have a random walk which we will start at an interior vertex and kill at the first boundary vertex. The transition probabilities $P^A(x, y)$ are equal to $P(x, y)$ for $x, y \in \Gamma \cap A$. If $x \in \Gamma \cap A$ is adjacent with one boundary point $w_\alpha = (\alpha, x, y) \in \partial A$ then the interval (x, y) is splitted into two parts and the transition probability modified accordingly:

$$P^A(x, w_\alpha) = \frac{\frac{1}{\alpha} \mu_{x,y}}{\sum_{z \neq y} \mu_{x,z} + \frac{1}{\alpha} \mu_{x,y}},$$

In other words new points are introduced as edge splitting points on the boundary of A and the walk is defined inside A as usual, choosing a neighbor proportional to the conductance. On vertices next to the boundary the walk tends to choose short edges with small α, which get bigger weights by $1/\alpha$. The walk is killed as soon as it reaches a boundary point. The exit time of the random walk is

$$T_A = \min \{n : X_n \in \Gamma \backslash A\},$$

and the mean exit time for the walk starts in $x \in \Gamma$ is defined as

$$E_x(A) = \mathbb{E}^A (T_A | X_0 = x),$$

where $\mathbb{E}^A$ is the expected value with respect to the probability measure P^A induced by the random walk X_n starting at $X_0 = x \in \Gamma$ and killed when it leaves A.

Remark 1 The notions of weight, capacity and resistance need a bit of explanation. Capacity is the reciprocal of resistance, shorter subintervals have smaller resistance and bigger capacity, while the weight assigned to the subinterval is proportional to its length. In that sense weight and capacity are not the same on subintervals while they numerically coincide on full intervals.

One can assume that the resistance is not uniform along the edges but there is a resistivity $\rho(s)$ along it and

$$R_{w_0,w_\alpha} = \int_0^\alpha \rho(s)\, ds.$$

This extension is not discussed here, but seems tractable and the whole machinery can be generalized to it without essential change.

Problem 1 (Maximal exit time) Let $x \in \Gamma$ and $M > 0$ be given. Find a set $F \ni x$, $F \subset W$ with volume $\mu(F) \le M$ and maximal expected exit time $E_x(F)$.

Problem 2 (Minimal capacity) Let $F \subset W$ be a fixed set and $M > 0$ be given. Let $Cap(D, F)$ denote the capacity of the 'annulus' $F \backslash D$ for $D \subset F$, more precisely

$$Cap(D, F) = \inf_{f \in H} \sum_{w,w' \in F \cup \partial F \cup \partial D} \left(f(w) - f(w')\right)^2 \mu_{w,w'},$$

where $H = H(D, F)$ is the set of functions $f : W \to \mathbb{R}$, $f|_{\overline{D}} \ge 1$ and $f|_{\overline{\Gamma \backslash F}} = 0$. Here again the boundary crossing edges are splitted and only the parts in $\overline{F \backslash D}$ is considered. We seek for a set D such that $D \subset F$, $\mu(D) \le M$ and the capacity $Cap(D, F)$ is minimal.

3 The Berman-Konsowa Principle

The other model that we will use is the *path system* of the graph. Consider a pair of sets (D, F), where $D \subset F$. Denote $L = L_{D,F}$ the set of all finite paths connecting $\partial_i D$ and $\partial_o F$ cropped at the boundary of the sets. Denote the ends of a path $l \in L_{D,F}$ by d_l and z_l, respectively. The path-graph on $F \backslash \partial_i D$ will be defined between y's and $\Gamma \backslash F$ and completed with common, unsplitted edges (d, y) reaching ∂D, see Fig. 1. (If $y \in \partial_0 D$ but $y \notin F$ we consider the single edge (d, y) as a path.)

We introduce $\mathcal{P}_{D,F}$ as the set of all probability measures on $L_{D,F}$, and let $Q_{D,F}$ be the edge set induced on $\overline{F} \backslash D$ by the original graph.

Definition 3.1 A flow between D and $\Gamma \backslash F$ is a function on $Q_{D,F}$. A flow function Φ is nonnegative and satisfies the following rules.

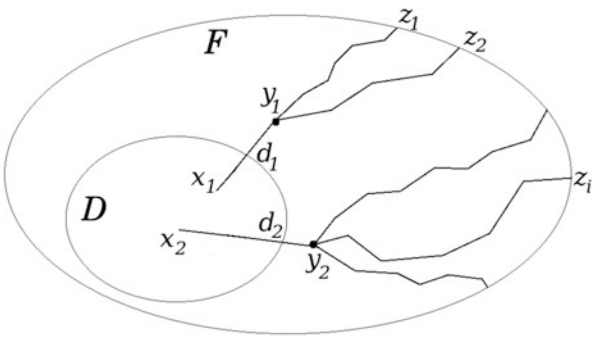


Fig. 1 Paths starting inside $F \setminus D$

1. $\Phi(x,y)\,\Phi(y,x) = 0 \quad \forall (x,y) \in Q_{D,F}$, i.e. the flow is one-directional,
2. for $x \in F \backslash D$

$$\sum_{y:(x,y)\in Q_{D,F}} \Phi(x,y) = 0,$$

3.

$$\sum_{d\in D, y:(d,y)\in Q_{D,F}} \Phi(d,y) = \sum_{z\in\partial F, y:(y,z)\in Q_{D,F}} \Phi(y,z)\,, \tag{3.1}$$

4. $\Phi(x,d) = \Phi(z,y) = 0$ for all $x, y \in F\backslash D, d \in D, z \in \partial F$. In addition we say that Φ is a unit flow if $\sum_{d\in D, y:(d,y)\in Q_{D,F}} \Phi(d,y) = 1$.

We define a new network (Γ_L, Q_L, μ_L) based on the path system $L = L_{D,F}$. That will be the set of paths connecting D and $\Gamma\backslash F$ with vertex and edge replicas of the original graph, to ensure that the path have no common vertices except at their endpoints. The objects of the new graph will be labeled by $l \in L$. Each $l \in L$ is a sequence of edges. We redefine the vertex set. For each $x \in F\backslash D$ let x_l be a vertex in Γ_L if $x \in l \cap (F\backslash D)$, formally: $\Gamma_L = \{x_l : x \in F\backslash D \cap l \text{ and } l \in L\}$. Edges are kept along the paths. We associate a new resistance $R^l_{x,y}$ to each edge on l with respect to a probability measure $\mathbb{P} \in \mathcal{P}_{D,F}$. If $(x,y) \in Q_{D,F}$ the flow can be decomposed into separate flows along disjoint paths

$$\Phi_{\mathbb{P}}(x,y) = \sum_{l':l'\ni(x,y)} \mathbb{P}\left(l'\right)$$

$$\mu^l_{x,y} = \mu_{x,y} \frac{\mathbb{P}(l)}{\Phi_{\mathbb{P}}(x,y)}$$

$$R^l_{x,y} = \left(\mu^l_{x,y}\right)^{-1} = R_{x,y} \frac{\Phi_{\mathbb{P}}(x,y)}{\mathbb{P}(l)}. \tag{3.2}$$

The path l has resistance $r_l = \sum_{(x,y)\in l} R^l_{x,y}$ and its conductance is $\mathrm{Cap}^{\mathbb{P}}(l) = 1/r_l$. Finally the capacity or conductance determined by $\mathbb{P}$ between D and ∂F is

$$\mathrm{Cap}^{\mathbb{P}}(D,F) = \sum_{l\in L_{D,F}} \mathrm{Cap}^{\mathbb{P}}(l).$$

Remark 2 Let us observe that the edge weights are shared between the paths, it is contained proportional to the probability measure. For each edge

$$\mu^l_{x,y} = \mu_{x,y} \frac{\mathbb{P}(l)}{\sum_{l':l'\ni(x,y)} \mathbb{P}(l')},$$

and consequently

$$\sum_{l:l\ni(x,y)} \mu^l_{x,y} = \mu_{x,y}.$$

Theorem 3.2 (Berman-Konsowa principle)

$$\mathrm{Cap}(D,F) = \max_{\mathbb{P}\in\mathcal{P}_{D,F}} \mathrm{Cap}^{\mathbb{P}}(D,F).$$

In what follows this nice path system will play a particular role. Let us mention that the capacity potential defines an important unit flow which minimizes the energy dissipation of the network. Let $\tau_C = \min\{k : X_k \in C\}$ be the hitting time of the set C and $v(y) = P\left(\tau_D < \tau_{\Gamma\setminus F} | X_0 = y\right)$. The natural flow generated by the properly adjusted external source is

$$\Phi(x,y) = (v(x) - v(y))_+ \mu_{x,y}, \tag{3.3}$$

where $a_+ = \max\{a, 0\}$.

4 Sets with Minimal Capacity

Let D be an optimal solution of Problem 2. Then we may assume that for all $w \in \partial D$, $w = (\alpha, x, y)$ with $\alpha \in (0,1)$, i.e., the boundary points of D are internal points of edges. We can assume even more, that there is a small $\varepsilon > 0$ such that

$$\alpha \in (\varepsilon, 1-\varepsilon).$$

If it is not the case, given that ∂D is finite, with an arbitrary small change of the volume M that can be ensured.

Consider the Berman-Konsowa path system $L_{\partial_o D,F}$ and let us extend each path l which connect $y_l \in \partial_o D$ to F with the edge segment (d_l, y_l), where $d_l \in \partial D$ and $d_l = (\alpha_l, x_l, y_l)$ for some $0 \le \alpha_l \le 1$. Then the resistance from d_l can be calculated as follows. We have

$$\text{Cap}_L(y_l, \partial F) = \sum_{\tilde{l}: y_l \in \tilde{l}} \text{Cap}\left(y_l, f_{\tilde{l}}\right),$$

where $f_l \in \partial F \cap l$ and $R_L(y_l, \partial F) = 1/\text{Cap}_L(y_l, \partial F)$

$$R_L(d_l, \partial F) = R_{d_l, y_l} + R_L(y_l, \partial F).$$

Finally, $\text{Cap}_L(d_l, \partial F) = 1/R_L(d_l, \partial F)$ and we have

$$\text{Cap}(D, F) = \sum_{d \in \partial D} \text{Cap}_L(d, \partial F).$$

Let us recognize, that the path system we have used here is smaller than the path system in the original B-K construction, since the border crossing edges are not split. For that reason we will refer to this construction as reduced B-K path system.

In order to investigate the optimal set of Problem 2 we use the Lagrange method and consider small perturbations of the optimal set. Let us consider a function $\xi : cD \to (0, 1)$ which defines the boundary of the set D_ξ with $w = (\xi(x, y), x, y) \in \partial D_\xi$.

We consider the reduced B-K path system over (D, F) and fix the resistances $R^l_{x,y}$ defined in (3.2) by the capacity potential and optimal flow (3.3) . We reserve $\mathbb{P}$ for the optimal distribution and $\tilde{\mathbb{P}}$ will denote an arbitrary other one on the fixed set of paths L. We shall consider in many cases a fixed set of paths L with different weights, in that case the resistances, conductances on the path system with respect to the probability $\mathbb{P}, \tilde{\mathbb{P}}$ will be denoted by R, Cap, and $\tilde{R} = \tilde{R}_L = R^{\tilde{\mathbb{P}}}_L$, $\widetilde{\text{Cap}} = \widetilde{\text{Cap}}_L = \text{Cap}^{\tilde{\mathbb{P}}}_L$, respectively. We shall drop the sub and superscripts if it does not cause ambiguity.

Remark 3 The Berman-Konsowa principle says that for any set $D \subset F$ and any weight system $\tilde{\mathbb{P}}$ with the corresponding $\widetilde{\text{Cap}}_L$

$$\text{Cap}(D, F) = \max_{\mathbb{P}} \text{Cap}^{\mathbb{P}}_L(D, F) \geq \widetilde{\text{Cap}}_L(D, F).$$

In particular if D is optimal, and $\tilde{D}$ is another set in F then

$$\text{Cap}\left(\tilde{D}, F\right) \geq \text{Cap}(D, F) \geq \widetilde{\text{Cap}}_L(D, F).$$

Lemma 4.1 *If $D \subset F, \mu(D) \leq M$ minimizes the capacity on the path system $L_{D,F}$ (with weights defined by the optimal $\mathbb{P}$) then D is optimal for Problem 2.*

Proof Let $\tilde{D} \subset F, \mu\left(\tilde{D}\right) \leq M$ be another set and $\tilde{L}$ be the path system defined by $(\tilde{D}, F)$. Then from

$$\text{Cap}^{\mathbb{P}}_L\left(\tilde{D}, F\right) \geq \text{Cap}^{\mathbb{P}}_L(D, F)$$

and from the capacity definition and Remark 3, we have the statement:

$$\mathrm{Cap}\left(\tilde{D},F\right)=\widetilde{\mathrm{Cap}}_{\tilde{L}}\left(\tilde{D},F\right)\geq \mathrm{Cap}_L^{\mathbb{P}}\left(\tilde{D},F\right)\geq \mathrm{Cap}_L^{\mathbb{P}}\left(D,F\right)=\mathrm{Cap}\left(D,F\right).$$

□

For each path $l\in L$, we introduce the resistance r_l and the weight μ_l of the whole path:

$$r_l=\sum_{(z,v)\in l}R_{z,v}^l \quad \text{and} \quad \mu_l=\sum_{(z,v)\in l}\mu_{z,v}^l,$$

where $(z,v)\in l$ are the edges of path l.

The proof of the following statement is given in the Appendix.

Proposition 4.2 *If D is optimal, then in the path system for each $l\in L_{D,F}$*

$$\mu^l\left(x_l,y_l\right)r_l=\text{const}, \tag{4.1}$$

where $(x_l,y_l)\in cD$ denotes the crossing edge of l.

Remark 4

1. The simplest case is to fix $x\in\Gamma$ and look for a set D with $x\in D\subset F$, $\mu\left(D\right)\leq M$ which minimizes $\mathrm{Cap}\left(x,D\right)$.
2. The following example shows that an optimal set D need not be unique. Consider two copies of $L_i=\{0_i,1_i,2_i,3_i,4_i,5_i\}$, $i=1,2$ with edges between direct neighbours and join them by setting $0_1=0_2$. We switch to continuous setup. Let

$$m_i\left(s\right)=\begin{cases}2 \text{ if } s\in[0,1]\\ 4 \text{ if } s\in(1,2]\\ 3 \text{ if } s\in(2,5]\end{cases}$$

be the mass density along L_i. Denote the mass and resistance of the ray from 0 to a point $x\in[0,5]$ by m and r. Then

$$m(x)=\begin{cases}2x & \text{if } 0\leq x<1\\ 2+4\left(x-1\right) & \text{if } 1<x\leq 2\\ 6+3\left(x-2\right) & \text{if } 2<x\leq 5\end{cases}$$

$$r(x)=\begin{cases}\frac{x}{2} & \text{if } 0\leq x<1\\ 1/2+\frac{1}{4}\left(x-1\right) & \text{if } 1<x\leq 2\\ \frac{3}{4}+\frac{1}{3}\left(x-2\right) & \text{if } 2<x\leq 5\end{cases}.$$

For the calculation of capacity we pick a point x on the first ray for which $m(x)$ is the total mass from 0 to it. We allocate the rest of the mass to the point y on the second ray, that is $m(y) = 12 - m(x)$ if $M = 12$. We define the inverse function p of the function m:

$$p(t) = \begin{cases} t/2 & \text{if } 0 \le t < 2 \\ \frac{1}{4}(t+2) & \text{if } 2 < t \le 6 \\ t/3 & \text{if } 6 < t \le 15 \end{cases}.$$

The capacity is $\mathrm{Cap}(x, D) = g(t) = 1/r(p(t)) + 1/r(p(12-t))$ expressed in the mass t used on the first ray, namely $t = m(x)$. It is easy to see, that this function has two global minimal solutions.

3. This example shows that no unique optimal solution can be guaranteed. The last resort is provided by the observation made in Lemma 4.1. If a set is optimal on its own B-K path system, it can be found by the Lagrange method, and it is optimal with respect to the original problem. Also let us recall Remark 3 which helps to sort out non-optimal sets. Since if we find a weight system $\mathbb{P}'$ over L on which the candidate set D^* is not optimal then it can not be optimal.

5 The Exit Time

We are going to find optimal sets which maximize $E_x(F)$ with $x \in F$, where $x \in \Gamma$ and $\mu(F) \le M$ fixed. In order to find an optimal set we try to maximize simultaneously $R(D_m, F)$ where D_m is the level set of the Green function $G^F(x, y)$ of volume m, $0 \le m \le M$. We defer the statement of the result to the end of this section to avoid repetition of technical notation. As in Sect. 4 we assume that the boundary points of the optimal set are inside the edges, i.e., ε-separated from the endpoints.

From now on we work on the reduced path system and weights are defined by the optimal flow. The path system is $\{l_i\}_{i=1}^n$, denote $z_i = l_i \cap \partial F$. Let

$$e_{l_i} = E_i(z_i) = \mathbb{E}(T_{z_i} | Z_0 = x)$$

the exit time on the path l_i of the random walk Z_n on l_i determined by the weights on l_i. Denote $m_i = m_{l_i}$ the volume of the path l_i and $R_i = R_{l_i}$ the resistance of it.

Definition 5.1 The local Green function (Green kernel) G^F $\left(g^F(x, y)\right)$ is defined by the transition probabilities $P_n^F(x, y)$ of the random walk, killed on exiting the set $F \subset \Gamma$ is the following:

$$G^F(x, y) = \sum_{n=1}^{\infty} P_n^F(x, y),$$

$$g^F(x,y) = G^F(x,y)/\mu(y).$$

In the following we summarize some known facts about the Green function and the exit time of random walks (for more details see [11]). It is know that

$$E_x(F) = \sum_{y \in F} g^F(x,y)\,\mu(y).$$

Furthermore, on the graph and on the cable system for any $w \in F$

$$G^F(x,w) = g^F(x,w)\,d\mu(w),$$
$$E_x(F) = \int_F g^F(x,w)\,d\mu(w), \tag{5.1}$$

where

$$g^F(x,w) = R(H_w, \partial F),$$

where $H_w = \{v : g^F(x,v) > g^F(x,w)\}$ is the super-level set with boundary of the equipotential surface B_w. (Let us remark here that g^F on the cable system is linear extension of g^F on the graph.) On the other hand we know that in the path decomposition we have for a given l_i that the Green kernel $g_i^{z_i}(x,w) = g_{l_i}^{z_i}(x,w) = R_i(w,z_i)$ and similarly to (5) we have that

$$e_{l_i} = \int_{l_i} g_i^{z_i}(x,w)\,d\mu(w) = \int_{l_i} R_i(w,z_i)\,d\mu(w),$$
$$\frac{R}{R_i} = \frac{R(x,F)}{R_i(x,z_i)} = \frac{R(B_w,\partial F)}{R_i(w,z_i)}$$

consequently for all path l_i

$$g^F(x,w) = R(B_w,F) = \frac{R}{R_i} r_i(w,z_i) = \frac{R}{R_i} g_i^{z_i}(x,w). \tag{5.2}$$

Since the path system splits each edge, we have

$$d\mu(w) = \sum_{i:\, w \in l_i} d\mu_i(w), \tag{5.3}$$

where in general $d\mu_i = P(l_i)\,d\mu$ and in particular $P(l_i) = \frac{R}{R_i}$. Here z_i's are not necessarily different. In the next step we shall join the paths which have common endpoints, i.e., the boundary crossing edge is shared by them.

$$E_x(F) = \int_F g^F(x,y)\,d\mu(y) = \sum_i \int_{l_i} g^F(x,y)\,d\mu_i(y) \tag{5.4}$$
$$= \sum_z \sum_{l_i \ni z} \int_{l_i} g^F(x,y)\,d\mu_i(y)$$

As earlier we should handle with care the paths ending at the same vertex (sharing a boundary crossing edge). The weights split on edges and hence the measure on vertexes add up as in (5.3)

$$d\mu_z(y) = \sum_{i:\, z \in B_y \cap l_i} d\mu_i(z),$$

while for $z \in B_y \cap l_i$

$$\begin{aligned} g^F(x,y) &= g^F(y,x) = \frac{R}{R_i} g^z(y,x) \\ &= \frac{R}{R_i} g_i^z(y,x) = \frac{R}{R_i} g_i^z(x,y). \end{aligned}$$

As a consequence of (5.4), (5.2) and the notation

$$e_l = \int_l g_l^z(x,y)\, d\mu_z(y),$$

we have that

$$E_x(F) = \sum_{z \in \partial F} \sum_{i: z \in l_i} \int_{l_i} g_{l_i}^F(x,y)\, d\mu_i(y) = \sum_l \frac{R}{R_l} e_l.$$

As a result we have the following observation.

Lemma 5.2 *For the set F the exit time $E_x(F)$ has the form*

$$E_x(F) = R \sum_l \frac{e_l}{R_l}.$$

We introduce the following notations: $C = \mathrm{Cap}(x,F) = 1/R(x,F)$

$$C_l = \frac{1}{R_l}, \quad \gamma_l = \frac{C_l}{C}, \quad \tilde{e} = \sum_l \gamma_l e_l, \quad \varphi_z = \frac{\delta_l e_l}{\tilde{e}},$$

where δ_z is such that $\mu_l R_l = (1+\delta_l)\, e_l$ holds.

Theorem 5.3 *If F is optimal for Problem 1, then it satisfies for all the B-K path l and its endpoint $z \in l \cap \partial F$ that*

$$\frac{R_l(x,z)\, \mu(z)}{(1+\varphi_l)^{1/2}} = \text{const}.$$

The proof is deferred to the Appendix.

As in case of the capacity problem, the obtained solution is not necessarily optimal or unique, since it is only a necessary and not a sufficient condition for optimality in general (see Remark 4 2. and 3.).

Appendix

Proof of Proposition 4.2 Let us recall that we assume that D is an optimal set and slightly change its boundary along the border crossing edges. We consider the Lagrange function with multiplier $\lambda \in \mathbb{R}$:

$$\mathrm{Cap}_L^{\mathbb{P}}\left(D_\xi, F\right) + \lambda\mu\left(D_\xi\right).$$

Denote $\xi_l = \xi_l(x, y) : w_l = (\xi_l, x, y) \in \partial D$ forming the perturbation vector $\xi = [\xi_l]$. Let $z_l = \partial F \cap l$ be the endpoint of the path l at the boundary of F.

$$\frac{\partial}{\partial \xi_l}\left[\sum_l \mathrm{Cap}_L^{\mathbb{P}}(w_l, z_l) + \lambda\mu\left(D_\xi\right)\right].$$

Setting the derivative zero and using $r_l = R^l(w_l, z_l)$ we have that

$$\begin{aligned} 0 &= \frac{\partial}{\partial \xi_l}\left[\sum_l \mathrm{Cap}_L^{\mathbb{P}}(w_l, z_l) + \lambda\mu\left(D_\xi\right)\right] = \\ &= \frac{\partial}{\partial \xi_l}\left[\frac{1}{r_l} + \lambda\mu_l\right] \\ &= -\frac{R^l(x_l, y_l)}{r_l^2} + \lambda\mu^l(x_l, y_l) \end{aligned}$$

for all path $l \in L$ and

$$\mu^l(x_l, y_l)\, r_l = \mathrm{const}$$

is a necessary condition for the optimality. □

Proof of Theorem 5.3 We consider the variational problem

$$\max_{F' :\, \mu(F') \le M} E_x\left(F'\right).$$

Assume that F is optimal with a path system L and the probability $\mathbb{P}$ on it. As in the case of the capacity we perturb F in a small neighborhood. The maximal solution should satisfy for a suitable λ and for all path l that

$$\frac{\partial}{\partial s_l}\left[E_x(F)+\lambda\mu(F)\right]=0$$

$$\frac{\partial}{\partial s_l}\left[R\sum_p\frac{e_p}{R_p}+\lambda\mu(F)\right]=\frac{\partial}{\partial s_l}\left[R\sum_p\left(\frac{e_p}{R_p}+\lambda\mu_p\right)\right]=0,$$

where s_l is the length of l and we use μ_l for the volume of the path l. Let $E=\sum\frac{e_l}{R_l}$, the density of μ is $\mu(z_l)=\frac{d\mu}{ds}|_{s_l}$, where s is the arc length parametrization of $l_z : w(s_l)=z_l$. Furthermore, $\mu_l=\mu(z_l)$ and the density of resistance is $\rho(z_l)=1/\mu(z_l)$, then the derivative is as follows

$$\frac{\partial}{\partial s_l}\left[R\sum_p\left(\frac{e_p}{R_p}+\lambda\mu_p\right)\right]=\left(\frac{\partial}{\partial s_l}R\right)E+R\frac{\partial}{\partial s_l}E+\lambda\mu(z_l).$$

One can find that

$$\left(\frac{\partial}{\partial s_l}R\right)=\frac{\partial}{\partial s_l}\frac{1}{\sum_p\frac{1}{R_p}}=R^2\frac{\rho(z_l)}{R_l^2}$$

and

$$\begin{aligned}\frac{\partial}{\partial s_l}e_l&=\frac{\partial}{\partial s_l}\int r(w_s,z_l)\,\mu(w(s))\,ds\\&=\frac{\partial}{\partial s_l}\int_0^{s_l}\int_s^{s_l}\rho(w(t))\,dt\mu(w(s))\,ds\\&=\rho(z_l)\,\mu_l,\end{aligned}$$

$$\frac{\partial}{\partial s_l}E=\frac{\partial}{\partial s_l}\frac{e_l}{R_l}=\rho(z_l)\,\frac{\mu_lR_l-e_l}{R_l^2}.$$

It is trivial that $e_l\leq\mu_lR_l$, so the defined δ_l is nonnegative. Furthermore,

$$\frac{\partial}{\partial s_l}E=\frac{\rho(z_l)}{R_l^2}\delta_le_l.$$

$$\frac{\partial}{\partial s_l}E_x(F)=R^2\frac{\rho(z_l)}{R_l^2}E+R\frac{\rho(z_l)}{R_l^2}\delta_le_l+\lambda\mu(z_l)=0$$

$$R\frac{\rho(z_l)^2}{R_l^2}E+\frac{\rho(z_l)^2}{R_l^2}\delta_le_l=\text{const.}$$

□

References

1. R. Bañuelos, Isoperimetric inequalities in probability, geometry, and PDE's, http://www.math.purdue.edu/~banuelos/btlecture.pdf
2. M. Barlow, J. Pitman, M. Yor, *On Walsh's Brownian Motions*. Lecture Notes in Mathematics, vol. 1372 (Springer, Berlin, 1989), pp. 275–293
3. M.T. Barlow, R.F. Bass, T. Kumagai, Stability of parabolic Harnack inequalities on metric measure spaces. J. Math. Soc. Jpn. **58**(2), 485–519 (2006)
4. K.A. Berman, M.H. Konsowa, Random paths and cuts, electrical networks, and reversible Markov chains. SIAM J. Discret. Math. **3**(3), 311–319 (1990)
5. T. Carroll, Old and new on the Bass note, the Torsion function and the hyperbolic metric. Irish Math. Soc. Bull. **47**, 41–65 (2001)
6. P.G. Doyle, J.L. Snell, *Random Walks and Electric Networks*. Carus Mathematical, Monographs, 22 (Mathematical Association of America, Washington, 1984)
7. A. Grigory'an, A. Telcs, Two-sided estimates of heat kernels on metric measure spaces. Ann. Probab. **40**(3), 1212–1284 (1990)
8. F.T. den Hollander, S. Jensen, Berman-Konsowa Principles for Reversible Markov Jump Processes (preprint). http://arxiv.org/abs/1309.1305
9. G. Pólya, G. Szegő, *Isoperimetric Inequalities in Mathematical Physics (No. 27)* (Princeton University Press, Princeton, 1951)
10. A. Telcs, Isoperimetric inequalities for Random Walks. Potential Anal. **19**(3), 237–249 (2003)
11. A. Telcs, *The Art of Random Walks*. Lecture Notes in Mathematics, vol. 1885 (Springer, Berlin/New York, 2006)