CONTEMPORARY MATHEMATICS

744

Dynamics: Topology and Numbers

Conference
Dynamics: Topology and Numbers
July 2–6, 2018
Max Planck Institute for Mathematics, Bonn, Germany

Pieter Moree
Anke Pohl
L'ubomír Snoha
Tom Ward
Editors

AMERICAN MATHEMATICAL SOCIETY

Dynamics: Topology and Numbers

CONTEMPORARY MATHEMATICS

744

Dynamics: Topology and Numbers

Conference
Dynamics: Topology and Numbers
July 2–6, 2018
Max Planck Institute for Mathematics, Bonn, Germany

Pieter Moree
Anke Pohl
L'ubomír Snoha
Tom Ward
Editors

AMS AMERICAN MATHEMATICAL SOCIETY

2010 *Mathematics Subject Classification.* Primary 11J70, 20F65, 22D40, 30E05, 37A15, 37A20, 37A30, 37B05, 54H20, 60B15.

Library of Congress Cataloging-in-Publication Data

Names: Moree, Pieter, 1965– editor.
Title: Dynamics: topology and numbers: Conference on dynamics: topology and numbers, July 2–6, 2018, Max Planck Institute for Mathematics, Bonn, Germany / Pieter Moree, Anke Pohl, L'ubomír Snoha, Tom Ward, editors.
Description: Providence, Rhode Island: American Mathematical Society, [2020] | Series: Contemporary mathematics, 0271-4132; volume 744 | This was the fourth conference at MPIM on the theme of dynamical systems and its relationships with the diverse fields of number theory, geometry, topology, ergodic theory, and combinatorics. | Includes bibliographical references.
Identifiers: LCCN 2019040097 | ISBN 9781470451004 (paperback) | ISBN 9781470454548 (ebook)
Subjects: LCSH: Dynamics–Congresses. | Number theory–Congresses. | Group theory–Congresses. | AMS: Number theory – Diophantine approximation, transcendental number theory [See also 11K60] – Continued fractions and generalizations [See also 11A55, 11K50]. | Group theory and generalizations – Special aspects of infinite or finite groups – Geometric group theory [See also 05C25, 20E08, 57Mxx]. | Topological groups, Lie groups {For transformation groups, see 54H15, 57Sxx, 58-XX. For abstract harmonic analysis, see 43-XX} – Locally compact groups and their algebras – Ergodic theory on groups | Functions of a complex variable {For analysis on manifolds, see 58-XX} – Miscellaneous topics of analysis in the complex domain – Moment problems, interpolation problems. | Dynamical systems and ergodic theory [See also 26A18, 28Dxx, 34Cxx, 34Dxx, 35Bxx, 46Lxx, 58Jxx, 70-XX] – Ergodic theory [See also 28Dxx] – General groups of measure-preserving transformations [See m | Dynamical systems and ergodic theory [See also 26A18, 28Dxx, 34Cxx, 34Dxx, 35Bxx, 46Lxx, 58Jxx, 70-XX] – Ergodic theory [See also 28Dxx] – Orbit equivalence, cocycles, ergodic equivalence relations. | Dynamical systems and ergodic theory [See also 26A18, 28Dxx, 34Cxx, 34Dxx, 35Bxx, 46Lxx, 58Jxx, 70-XX] – Ergodic theory [See also 28Dxx] – Ergodic theorems, spectral theory, Markov operators For op | Dynamical systems and ergodic theory [See also 26A18, 28Dxx, 34Cxx, 34Dxx, 35Bxx, 46Lxx, 58Jxx, 70-XX] – Topological dynamics [See also 54H20] – Transformations and group actions with special proper | General topology {For the topology of manifolds of all dimensions, see 57Nxx} – Connections with other structures, applications – Topological dynamics [See also 28Dxx, 37Bxx]. | Probability theory and stochastic processes {For additional applications, see 11Kxx, 62-XX, 90-XX, 91-XX, 92-XX, 93-XX, 94-XX} – Probability theory on algebraic and topological structures – Probabil
Classification: LCC QA843 .D959 2020 | DDC 510–dc23
LC record available at https://lccn.loc.gov/2019040097
Contemporary Mathematics ISSN: 0271-4132 (print); ISSN: 1098-3627 (online)

DOI: https://doi.org/10.1090/conm/744

---

10 9 8 7 6 5 4 3 2 1   25 24 23 22 21 20

The editors and participants dedicate this volume to the memory of our dear friend and colleague Sergiĭ Kolyada (December 7th, 1957 – May 16th, 2018)

Contents

Preface

This volume flows from the activity Dynamics: Topology and Numbers, held at the Max Planck Institute for Mathematics (MPIM) in Bonn in July 2018. The activity brought together about 90 researchers from 25 different countries, along with many of the long-term visitors at the MPIM. This was the fourth conference at the MPIM on the theme of dynamical systems and its relationships with the diverse fields of number theory, geometry, topology, ergodic theory, and combinatorics.

Sergiĭ at the MPIM (with permission of the Max Planck Institute for Mathematics)

The four activities in the same broad area were instigated by Sergiĭ, starting with the highly successful Algebraic and Topological Dynamics in 2004, continuing with Dynamical Numbers in 2009 and Dynamics and Numbers in 2014, and ending with Dynamics: Topology and Numbers in 2018. All four of these gatherings highlighted Sergiĭ's ability to bring diverse mathematicians together and to organise conferences with great energy, friendliness, enthusiasm, and commitment. The success of this series of conferences also reflects the effort and skill of the staff at the MPIM, who found ways around every difficulty and offered unfailing support to the various organisers (comprising David Ellwood, Sergiĭ Kolyada, Yuri Manin, Martin Möller, Pieter Moree, Anke Pohl, Tom Ward, and Don Zagier in various combinations).

Tragically, Sergiĭ passed away less than two months before the 2018 conference was due to take place. Instead of a rather celebratory event, those assembled were able to share fond memories of Sergiĭ and reflect on his many contributions. We can only echo Mike Shub's words when another mathematician passed away far too early: "Don't forget to say that we all liked him".

The editors wish to record their thanks to the staff at the Max Planck Institute, and to the many researchers who participated, for all the efforts they made to make this a stimulating and intellectually rich event. The editors also thank Alexandre Kosyak for help in contacting mathematicians in Ukraine connected to Sergiĭ.

Pieter Moree, Bonn
Anke Pohl, Bremen
Ľubomír Snoha, Banská Bystrica
Tom Ward, Leeds

Contemporary Mathematics
Volume **744**, 2020
https://doi.org/10.1090/conm/744/14975

The life and mathematics of Sergiĭ Kolyada

L'ubomír Snoha

Sergiĭ Kolyada was a distinguished Ukrainian mathematician in the area of low-dimensional dynamical systems and topological dynamics. He studied dynamical systems of the form (X, f), given by a (usually compact metric) space X and a continuous map $f\colon X \to X$. If we were to describe his work in keywords, then the following are probably the most appropriate: zero Schwarzian, triangular map, ω-limit set, topological entropy, topological transitivity, minimal map, minimal set, chaos, sensitivity, functional envelope of a dynamical system, and dynamical compactness. He introduced, together with his co-authors, several new notions including the functional envelope of a dynamical system, Li–Yorke sensitivity, dynamical topology, and dynamical compactness.

Sergiĭ's name appears in many variants.

Сергій Коляда (in his Ukrainian language publications);
Сергій Федорович Коляда (in the Ukrainian Wikipedia);
Сергей Федорович Коляда (in older publications in Russian);
Sergiĭ Kolyada (on his home page, most of recent papers, and his preferred version in the last few years);
Sergiĭ Kolyada (in some recent papers);
Sergii Koliada (in his last passport);
Sergiy Kolyada (on social media and in his Google Scholar profile);
also as Sergii Kolyada, Sergei Kolyada, Sergey Kolyada, S. F. Koljada and perhaps others.

Коляда in Ukrainian means "a Christmas carol".

1. On the life of Sergiĭ Kolyada

Sergiĭ Kolyada was born on December 7th, 1957 in Kolyady, a small village in the Shyshaky raion (district) in Poltava oblast (province, or region) in Ukraine (at that time, part of the Soviet Union). Kolyady is located between Kiev and Kharkiv, some 300km from Kiev. He and his younger brother attended school in the nearby village of Pryshyb. Because his mathematics teacher conducted classes

2010 *Mathematics Subject Classification.* Primary 01A70.
Key words and phrases. Biography, obituary, bibliography.
The author was supported by the Slovak Research and Development Agency under the contract No. APVV-15-0439. He is very grateful to Tom Ward for many improvements of an earlier version of the manuscript.

with him separately, he won the district mathematics Olympiad in the 7th grade. At the same time, a telegram was sent to the school that the Ukrainian Physics and Mathematics Lyceum, a boarding high school for talented pupils, located in Kiev and affiliated with Taras Shevchenko National University could be applied for. The family allowed him to apply and, having passed the entrance examinations, he became a student of this boarding school in 1972. From 1975 he studied at Taras Shevchenko University in Kiev, where he graduated with an M.Sc. in Mathematics in 1980.

Sergiĭ married Maria Vakaryuk in 1980, and they went on to have two daughters, Irina (who has two sons Artem and Maxim) and Natasha. His devotion to the family and the great pleasure he took in them was evident to all who knew him throughout his life.

After graduation he worked as a programming engineer, and it was only in 1985 that he started to work at the Academy of Sciences of the Ukrainian Soviet Socialist Republic (SSR)[1] in Kiev (first at the Institute of Hydrobiology and, from 1987, at the Institute of Mathematics). In 1987 he received the title of Candidate of Physical and Mathematical Sciences (C.Sc.), a post-graduate scientific degree corresponding to the PhD, when he defended his dissertation "Discrete dynamical systems with negative and zero Schwarzian derivative" under the supervision of Alexander N. Sharkovsky (Šarkovskiĭ).

Sergiĭ started what became a long and significant cooperation with the Max Planck Institute for Mathematics (MPIM) in Bonn in 2001. He co-organized several activities there in the area of dynamical systems and some of its near relatives. For more details about these activities, we refer to the article "Sergiy and the MPIM" by Pieter Moree in this volume.

In 2005 he received the title of Doctor of Physical and Mathematical Sciences (D.Sc.) after defending his dissertation "Topological dynamics: minimality, entropy and chaos" at the National Academy of Sciences of Ukraine in Kiev.

In 2006 he started to teach at the Taras Shevchenko University on a part-time basis, continuing his work at the Institute of Mathematics of the Academy of Sciences.

In 2010 the Ukrainian State Prize in Science and Technology was awarded to a group of researchers including Sergiĭ for a collection of publications entitled "Dynamical systems theory: modern methods and their applications".

Sergiĭ spent the summer semester of 2013 at the Technische Universität München as a John von Neumann Visiting Professor.

De facto in November 2017, and *de jure* a few months later, Sergiĭ became the head of the Department of Dynamical Systems and Fractal Analysis of the Institute of Mathematics, and he fulfilled these duties until his untimely death.

Sergiĭ was an exceptional organizer of a wide range of academic activities. He was a member of many organizing committees, and was highly successful in winning international grants for organizing conferences. He organized the Soviet–Spanish–Czechoslovakian Symposium "Dynamical Systems and their Applications" in Kiev (June 28–July 4, 1991) and the conference "Dynamical Systems and Ergodic Theory" in Katsiveli (August 21–30, 2000). He also co-organized activities and

[1]From 1991 to 1993 it was called the Academy of Sciences of Ukraine, and since 1994 its name has been the National Academy of Sciences of Ukraine.

conferences in the area of dynamical systems at the Max Planck Insititute in Bonn in 2004, 2006, 2009 and 2014; for more details see the article by Pieter Moree.

Being both very friendly, international in outlook, and naturally communicative, Sergiĭ had co-authors from all over the world, including Chile, China, France, Israel, the Netherlands, Slovakia, Spain, Ukraine and the U.S.A. He made research visits to institutions in Australia, Chile, China, Czech Republic, France, Germany, Hongkong, Slovakia, Spain and the U.S.A.

His public activities were numerous and impressive. He was a member of the American Mathematical Society, the Society of Friends of the Institut des Hautes Études Scientifiques, the Ukrainian Scientific Club, and the Kiev Mathematical Society. He played a decisive role in the relaunch of the Kiev Mathematical Society in 2006, and he served as President of the society for the years 2006–2014, and as Vice-President thereafter. Being a Ukrainian patriot, he did not miss any opportunity to support mathematical research in Ukraine and especially to encourage and sponsor young Ukrainian mathematicians and students of mathematics. He brought several talented Ukrainian students to MPIM, and suggested them for PhD positions there; for details see the article by Pieter Moree. Sergiĭ was co-chairman of the Competition of "Shevchenko Scientific Society in the US" and the "US-Ukraine Foundation" for young mathematicians in Ukraine. He initiated the creation of "M. Ostrogradskiy scholarships" from the all-Ukrainian charitable organization "Fund for the Promotion of Mathematical Science" to the ten best students of mathematics at Universities in Kiev, and in the years 2001-2003 he organized lecture series from different branches of mathematics for these students.

Sergiĭ's hobbies largely revolved around sports. As a student he played basketball and football for the teams of Taras Shevchenko University. While he later quit basketball, he remained faithful to football throughout his life. From 2006 he was a member of the 'old men' or 'veteran' football clubs Fortuna (in Kiev) and SFC Hofgarten United (in Bonn). Besides playing, he was also a knowledgeable fan of football more generally. Continuing this enjoyment of skilled physical activity, his younger daughter Natasha is a professional ballroom dancer, and Sergiĭ enjoyed watching the dance competitions she and her partner attended. He was always very proud of her international successes.

Sergiĭ passed away prematurely in Kiev on May 16th, 2018. He is missed, not only by his co-authors and by the community of dynamists as an excellent mathematician and a colleague always interested in the work of others, but also by all who met him. All those who knew him will miss his smile, his great kindness, and his willingness to help.

2. On the mathematical work of Sergiĭ Kolyada

We cannot discuss all of Sergiĭ's results, concentrating instead on those which he considered to be most important or that he liked most. Sergiĭ appreciated the special role played by examples in dynamical systems, and was particularly talented at constructing revealing examples and counter-examples.

We divide his mathematical work into themes. Some of his papers and results should, or at least could, be included into several themes, but we will mention them only within one. To keep this article to a reasonable length, we will only occasionally mention papers by other authors which are based on work by Sergiĭ.

Sergiĭ's smile will be missed by all who knew him
(photo with permission of Irina Kolyada)

2.1. Interval dynamics. Sergiĭ did his PhD, and wrote his first articles, in interval dynamics [**K1**, **K2**, **K3**, **K4**, **K5**, **K7**, **K8**]. Some of those results are also included in the monograph [**K51**], whose slightly expanded English version appeared as [**K52**]. In interval dynamics he studied the measure of repellers and quasi-attractors, the dynamics of one-parameter families of continuous maps and the dynamics of continuous maps with Schwarzian derivative of constant sign.

For instance, in [**K4**] he proved that unimodal maps with zero Schwarzian derivative have at most one attracting cycle inside the interval (by work of Singer, such a property was known for unimodal maps with negative Schwarzian [**Sin78**]). In [**K8**] he gave a new example of a one-parameter family λf of unimodal maps with negative Schwarzian, for which the monotonicity of bifurcations for periodic orbits does not hold (and the topological entropy is not monotone); the first such example appeared in [**Zdu84**].

2.2. Triangular maps. In the study of topological dynamics of triangular maps, Sergiĭ's name is among the first which should be mentioned. A *triangular map* is a continuous map $F\colon I^2 \to I^2$, where $I = [0,1]$, of the form

$$F(x, y) = (f(x), g(x, y)).$$

Such a map preserves 'vertical' fibres: the fibre over x is mapped to a fibre over $f(x)$. This is a special type of skew product. Triangular maps became popular in 1979, when Kloeden [**Klo79**] proved that Sharkovsky's theorem held for them. Thus it was natural to believe that in fact most of the results of interval dynamics could be carried over to them (to establish this is sometimes called Sharkovsky's program; for the state of the art of this program see the recent survey [**Šte16**]). Kolyada was the first to show that this belief was false. In [**K6**, **K9**, **K10**, **K12**, **K13**, **K15**, **K19**] he developed the basics of the theory of triangular maps; a reader interested in this topic should first of all read [**K13**]. In particular, he constructed an important and surprising example of a triangular map of type 2^∞ with positive topological entropy [**K9**, **K13**] (recall that interval maps of type 2^∞ have entropy zero; a map is of type 2^∞ if it has periodic orbits of periods all powers of two, and of no other periods).

Triangular maps also appear in some other papers, including in particular the paper [**K22**] which is discussed in Subsection 2.6.

2.3. ω-limit sets. Given a point x in a dynamical system (X, f), its ω-limit set $\omega_f(x)$ is the set of all limit points of the trajectory $x, f(x), f^2(x), \dots$. The topological structure of ω-limit sets depends on the phase space X. As established in [**ABCP89**], a non-empty closed subset M of $I = [0, 1]$ is an ω-limit set for some continuous map $f\colon I \to I$ if and only if M is either nowhere dense or is a union of finitely many non-degenerate closed intervals. If M is a subset of a vertical fibre in the square I^2, it can be an ω-limit set of a triangular map F of the square even if the topological structure of M is more complicated. The reason is that in this case the point x with $\omega_F(x) = M$ can be chosen outside that fibre. A full topological characterization of ω-limit sets of triangular maps which lie in just one vertical fibre, was found, in co-authorship, by Kolyada in [**K11**, **K14**]. The result has an interesting consequence we are going to describe.

By work of Dowker and Friedlander [**DF54**] for homeomorphisms, and of Sharkovsky [**Šar65**] for continuous maps, it is known that for dynamical systems on compact metric spaces the following are equivalent:

(1) (X, f) can be embedded as an ω-limit set in some larger system (Y, g),
(2) (X, f) is *f-connected* or *weakly incompressible* (meaning that there is no non-empty, proper, closed set $A \subsetneq X$ such that $f(A) \subseteq \operatorname{Int} A$).

The result from [**K11**, **K14**] mentioned above implies the following fact explicitly mentioned in [**K52**, p.21]:

A non-empty closed subset X of the unit interval is an ω-limit set (that is, there is a dynamical system (Y, g) containing X as an ω-limit set or, equivalently, X admits a continuous self-map f such that (X, f) is f-connected) if and only if X is *not* a disjoint union of a finite number of non-degenerate closed intervals and a non-empty countable set whose distance from at least one of those intervals is positive.

For some other results of Sergiĭ on ω-limit sets see [**K16**, **K17**, **K38**]. His last paper [**K50**] deals with special α-limit sets.

2.4. Topological entropy. Kolyada is a co-author of the extension of the notion of topological entropy to *non-autonomous* systems given by a compact metric space and a sequence of continuous self-maps of it:

$$X \xrightarrow{f_1} X \xrightarrow{f_2} X \xrightarrow{f_3} \cdots .$$

This was motivated by the wish to understand better the entropy of triangular maps, and in particular Bowen's formula [**Bow71**] giving an estimate for it. Basic properties of the entropy of non-autonomous systems have been proved, and basic counter-examples were published in [**K18**]. As an easy corollary of some results on the entropy of *non-autonomous* systems, the commutativity of the entropy of *autonomous* systems was obtained:

$$h(f \circ g) = h(g \circ f).$$

S. Kolyada, M. Misiurewicz and L'. Snoha in Banská Bystrica in June 2017 (working on [**K50**]; photo with permission of Štefan Gyürki)

Later it turned out that this formula was not new, see [**DM86**]. However, it was unknown among dynamists and probably it would still be forgotten if [**K18**] had not appeared. This is one of Sergiĭ's most cited papers.

In [**K23**], entropy for non-autonomous systems given by a sequence of piecewise monotone interval maps

$$I_1 \xrightarrow{f_1} I_2 \xrightarrow{f_2} I_3 \xrightarrow{f_3} \cdots$$

is studied (notice that the intervals are not fixed here) and it is proved that, under some additional assumptions, the Misiurewicz–Szlenk formula [**MS80**] for the topological entropy of piecewise monotone interval maps holds. In the autonomous case, the formula enables one to compute the entropy by counting the number of 'laps' of the iterates of the map. In the non-autonomous case, one has to count the number of laps of the compositions $f_n \circ \dots \circ f_2 \circ f_1$; the formula now works only under additional assumptions. As an application, the version of the Misiurewicz–Szlenk formula for non-autonomous systems allowed a dramatic simplification of the proof that Kolyada's old example of a triangular map of type 2^∞ from [**K9**, **K13**] has positive entropy.

Other papers of Sergiĭ dealing with topological entropy are [**K15**, **K22**, **K25**, **K28**, **K31**, **K32**, **K37**, **K42**].

2.5. Minimality. Minimality was one of Sergiĭ's favorite topics, and he contributed to this area significantly. For a survey on minimality we refer to [**K36**].

2.5.1. *Properties of minimal maps.* In topological dynamics, the most fundamental dynamical systems are the minimal ones. These are systems which have no non-trivial subsystems. More precisely, a dynamical system (X, f) defined by a topological space X and a continuous self-map f is called *minimal* if X does not contain any non-empty, proper, closed f-invariant set (a set $M \subseteq X$ is f-invariant if $f(M) \subseteq M$). In such a case we also say that the map f itself is minimal (note that here we consider only continuous maps). An equivalent definition is: (X, f) is minimal if for every $x \in X$, the orbit $\{x, f(x), f^2(x), \dots\}$ is dense in X.

Sergiĭ, with co-authors, showed that if (X, f) is minimal with X a compact metric space, then in many respects the continuous map f behaves like a homeomorphism [**K24**]:

- there is no non-empty redundant open set for f ($G \subseteq X$ is said to be *redundant* for f if $f(X \setminus G) = f(X)$);
- f is feebly open (i.e. it sends non-empty open sets to sets with non-empty interior);
- f preserves the topological size of a set in both directions. More precisely, the implication

$$A \subseteq X \text{ has property } P \Rightarrow \text{both } f(A) \text{ and } f^{-1}(A) \text{ have property } P$$

  holds for P meaning 'is nowhere dense', 'is dense', 'is of 1st category', 'is of 2nd category', 'is residual', 'has the Baire property', 'has non-empty interior';
- f is almost 1-to-1, meaning that the set $\{x \in X \mid \operatorname{card} f^{-1}(x) = 1\}$ is a dense G_δ in X.

Further, in [**K24**] the existence of minimal non-invertible maps on the torus was established (the existence of minimal homeomorphisms on the torus is well-known). This was the first example of a minimal non-invertible map on a manifold.

2.5.2. *Spaces admitting minimal maps.* A long-standing open problem in topological dynamics is to classify compact metric spaces admitting minimal maps or minimal homeomorphisms.

In [**K26**], non-homogeneous metric continua admitting minimal maps but not admitting minimal homeomorphisms are constructed (for the circle the converse is well-known: it admits a minimal homeomorphism but does not admit a minimal non-invertible map). One curious result from [**K29**] is that the interval $[0, 1]$ admits a sequence of continuous self-maps converging to the identity and forming a minimal non-autonomous dynamical system on $[0, 1]$. Some other contributions are explicitly or implicitly contained in papers dealing with minimal sets in a given space.

2.5.3. *Minimal sets.* Given a dynamical system (X, f), a set $M \subseteq X$ is called a *minimal set* if it is non-empty, closed and f-invariant and no proper subset of M has these three properties. So, a non-empty closed set $M \subseteq X$ is a minimal set if and only if $(M, f|_M)$ is a minimal system. Thus a system (X, f) is minimal if and only if X is the (unique) minimal set in (X, f). The basic fact discovered by G. D. Birkhoff is that in any compact system (X, f) there are minimal sets.

Minimal sets are fundamental objects of study in topological dynamics. A natural (and open) problem is to describe the possible minimal sets for dynamical systems in a given space. Two major contributions of Sergiĭ and his co-authors to this problem are the following results.

The first result states that for minimal sets on 2-manifolds the dichotomy "nowhere dense or everything" holds. More precisely, in [**K34**] it is proved that on compact connected 2-manifolds with or without boundary, a minimal set either is the whole manifold or is nowhere dense.

The second contribution is a very detailed description of minimal sets of continuous fibre-preserving maps in graph bundles, see [**K41**]. The complete results are too technical to describe here, but one particular corollary says that the fibre-preserving maps in *tree* bundles have only nowhere dense minimal sets. In particular, if F is a continuous triangular map in the square I^2 and M is a minimal set of F, then M is nowhere dense in the space $pr_1(M) \times I$ (pr_1 denotes the projection onto the first factor; the nowhere density of M in the square I^2 is trivial, but here $pr_1(M) \times I$ is a very small subspace of the square, since $pr_1(M)$ is a Cantor set or a finite set). Moreover, either a typical fibre of M is a Cantor set, or there is a positive integer N such that a typical fibre of M has cardinality N.

2.6. Topological transitivity. Topological transitivity, an important notion in topological dynamics and one of the ingredients in several well-known definitions of chaos, appeared often in Sergiĭ's research papers. He also co-authored two surveys on this topic, see [**K21**, **K35**]. In [**K22**] the connection between transitivity, density of the set of periodic points, and topological entropy for low dimensional continuous maps are investigated. The paper deals with this problem in the case of the n–star and the circle among the one-dimensional spaces and in some higher dimensional spaces. Particular attention is paid to extensions of transitive maps from a compact metric space X to triangular maps in $X \times [0, 1]$ without increasing topological entropy and with transitivity preserved. An extension theorem is obtained, saying that this is always possible when the transitive map in X is non-minimal. Later, in [**K37**], this was proved without the assumption of non-minimality. An analogue of the extension theorem from [**K22**] was also proved for many other dynamical properties, see [**BS03**, **Dir08**]. As shown already in [**K22**], such extension theorems can be used to find, on a given space, the infimum/minimum of the topological entropy of continuous maps with given dynamical property (the 'entropy bounds problem'). Other papers co-authored by Sergiĭ, where transitivity plays a central role, are [**K40**, **K43**, **K45**].

2.7. Chaos. Recall that if $(X; d)$ is a metric space and $f \colon X \to X$ a continuous map, then the dynamical system (X, f) is called *Li–Yorke chaotic* if there exists an uncountable set $S \subseteq X$ such that for all $x, y \in S$, $x \neq y$, the pair (x, y) is proximal but not asymptotic. That is

$$\liminf_{n\to\infty} d(f^n(x), f^n(y)) = 0 \quad \text{and} \quad \limsup_{n\to\infty} d(f^n(x), f^n(y)) > 0$$

(such a set S is called a *scrambled set*). There was a long-standing open problem whether positive topological entropy implies Li–Yorke chaos. The following theorem proved by Kolyada and his co-authors in [**K25**], is of fundamental importance in chaos theory:

If $(X; d)$ is a compact metric space and a continuous map $f \colon X \to X$ has positive topological entropy, then it is Li–Yorke chaotic.

Another important contribution is [**K27**] where Kolyada and Akin introduced the following new notion of chaos, which combines the classical notion of sensitivity with a Li–Yorke version of chaos.

Let $(X; d)$ be a compact metric space and $f \colon X \to X$ a continuous map. Then the system (X, f) is said to be *Li–Yorke sensitive* if there exists $\varepsilon > 0$ such that every $x \in X$ is a limit of points $y \in X$ such that the pair (x, y) is proximal but not ε-asymptotic. That is,

$$\liminf_{n\to\infty} d(f^n(x), f^n(y)) = 0 \quad \text{and} \quad \limsup_{n\to\infty} d(f^n(x), f^n(y)) \geq \varepsilon.$$

In [**K27**] the authors analyzed properties of such systems and compared them with other ones. The notion turned out to be quite popular, and some of the open problems formulated in that paper have already been solved. Other papers of Sergiĭ dealing with chaos or related topics are [**K30**, **K39**, **K47**].

2.8. The functional envelope of a dynamical system. If (X, f) is a dynamical system given by a compact metric space $(X; d)$ and a continuous map $f : X \to X$ then the *functional envelope* of (X, f) is, according to [**K32**], the dynamical system $(S(X), F_f)$ whose phase space $S(X)$ is the space (in general a non-compact space) of all continuous self-maps of X with the compact-open topology, and the map $F_f \colon S(X) \to S(X)$ is defined by $F_f(\varphi) = f \circ \varphi$ for any $\varphi \in S(X)$. Since the metric space $(X; d)$ is compact, each of the following two metrics is compatible with the compact-open topology on $S(X)$:

- the metric d_U of uniform convergence: $d_U(\varphi, \psi) = \sup_{x \in X} d(\varphi(x), \psi(x))$;
- the Hausdorff metric d_H, derived from the metric $d_{\max}((x_1, y_1), (x_2, y_2)) = \max\{d(x_1, x_2),\, d(y_1, y_2)\}$ on $X \times X$, applied to pairs of maps (no distinction is made between a map and its graph).

Though the metrics d_U and d_H are equivalent, they are not in general uniformly equivalent. Therefore the "uniform" and the "Hausdorff" envelopes of a system may differ with respect to dynamical properties which depend on the metric (for example, the topological entropy in non-compact envelopes may differ in this way).

The functional envelope of a system always contains a copy of the original system (hence the 'envelope' terminology). The motivation for the study of dynamics in functional envelopes comes from semi-group theory, from the theory of functional difference equations (mainly from [**SMR93**]) and from dynamical systems theory, see [**K32**]. We present a few results on dynamics in functional envelopes.

Let P be a property a map from $S(X)$ may or may not have. It is said to be a *range down property* if

$$\operatorname{range}\theta \subseteq \operatorname{range}\varphi \Longrightarrow (\varphi \text{ has } P \Rightarrow \theta \text{ has } P).$$

It is proved in [**K32**] that the following are range down properties: compactness of the orbit closure, non-emptiness of the ω-limit set, recurrence, simultaneous compactness and minimality of the orbit closure.

In [**K32**] it is also proved that while dense orbits in functional envelopes may exist, for many compact metric spaces X there are no dense orbits in the functional envelope $(S(X), F_f)$. In particular, this is true if X is a manifold.

If f has zero topological entropy, then both the uniform and the Hausdorff functional envelope may have infinite entropy; examples can be found in [**K32**]. By [**Mat13**], if X is a tree and f has zero or positive entropy, then the Hausdorff functional envelope has zero or infinite entropy, respectively. By [**K42**], if X is a Peano continuum or a compact metric space with continuum many connected components, then the only possible values of the entropy of the functional envelope are zero and infinity (regardless of whether we consider the uniform or the Hausdorff envelope). For more facts on the topological entropy of functional envelopes see [**DST17**].

2.9. Dynamical topology. Let us recall that in *topological dynamics* one investigates dynamical properties that can be described in topological terms. For example, topological transitivity is a dynamical property of a map, and is defined in terms of behaviour of open sets under the iterates of the map.

In contrast with this, in [**K43**] and [**K45**] Kolyada and his co-authors introduced the notion of *dynamical topology.* This is the area where one investigates topological properties of spaces of maps that can be described in dynamical terms. For instance, one can ask what are the topological properties of the space of all transitive maps on a given space.

If I is a real compact interval, denote by $\mathcal{T}$ the space of all (continuous) transitive maps $I \to I$, with the C^0-metric. Further, let $\mathcal{T}_{\mathrm{PM}}$ or $\mathcal{T}_{\mathrm{PL}}$ be the subspaces of all piecewise monotone, or piecewise linear elements of $\mathcal{T}$, respectively. One of the results from [**K43**] says that the spaces $\mathcal{T}$, $\mathcal{T}_{\mathrm{PM}}$ and $\mathcal{T}_{\mathrm{PL}}$ are contractible (and hence arc-wise connected) and are locally arc-wise connected. The investigation of the topology of various spaces of transitive interval maps is continued in [**K45**]. It is shown there that some loops that are not contractible in some of those spaces, can be contracted in slightly larger spaces. The methods developed in these papers are used in the recent paper [**Fan19**].

2.10. Dynamical compactness. A family $\mathcal{F}$ of subsets of $\mathbb{Z}_+ = \{0, 1, 2, \dots\}$ is called a *Furstenberg family* if it is *hereditary upward* (that is, $F_1 \subseteq F_2$, $F_1 \in \mathcal{F}$ together imply that $F_2 \in \mathcal{F}$). If $\mathcal{F}$ is a Furstenberg family, its *dual family* $k\mathcal{F}$ is defined as the family of all subsets of $\mathbb{Z}_+$ which are large enough to ensure that they intersect every set from $\mathcal{F}$ non-trivially.

If (X, T) is a dynamical system, $x \in X$, $G \subseteq X$ and $F \subseteq \mathbb{Z}_+$, put

$$T^F x = \{T^i x \mid i \in F\}$$

and

$$n_T(x, G) = \{n \in \mathbb{Z}_+ \mid T^n x \in G\}.$$

Now let X be a compact metric space. Recall that if (X, T) is a dynamical system and $x \in X$, then the classical ω-limit set of x is

$$\begin{aligned}\omega(x) = &\bigcap_{n\in\mathbb{N}} \overline{\{T^k(x) \mid k \geq n\}} \\ =&\{z \in X \mid \text{ for every neighbourhood } G \text{ of } z,\ n_T(x, G) \text{ is infinite}\}.\end{aligned}$$

By changing the set of times appearing in this definition, one can define the *ω-limit set of x with respect to a Furstenberg family $\mathcal{F}$*:

$$\omega_{\mathcal{F}}(x) = \bigcap_{F\in\mathcal{F}} \overline{T^F x}.$$

It is easy to show that

$$\omega_{\mathcal{F}}(x) = \{z \in X \mid \text{ for every neighbourhood } G \text{ of } z,\ n_T(x, G) \in k\mathcal{F}\}.$$

So,

$$\omega(x) = \omega_{\mathcal{F}_{\mathrm{cof}}}(x)$$

where $\mathcal{F}_{\mathrm{cof}}$ is the Furstenberg family of all cofinite sets.

By our assumption that the metric space X is compact, we have that $\omega(x) \neq \emptyset$ for every $x \in X$. Kolyada and his co-authors, see [**K44**], call a dynamical system (X, T) *dynamically compact with respect to* $\mathcal{F}$, if $\omega_{\mathcal{F}}(x) \neq \emptyset$ for every $x \in X$. They study this notion in [**K44**] and [**K48**]. In their considerations, the Furstenberg family $\mathcal{F}$ is a family with some dynamical meaning. Therefore they use terms such as *transitive compactness* or *sensitive compactness.* They investigate many classical notions of topological dynamics by using the concept of dynamical compactness. One remarkable fact proved in [**K48**] says that all dynamical systems are dynamically compact with respect to a Furstenberg family if and only if this family has the finite intersection property.

Sergiĭ had many mathematical plans for the future. It is our loss that he will not fulfil them.

Papers by Sergiĭ Kolyada

[K1] S. F. Kolyada and A. N. Sharkovskiĭ, *The measure of repellers of one-dimensional smooth mappings* (Russian), Functional-differential and difference equations, Akad. Nauk Ukrain. SSR, Inst. Mat., Kiev, 1981, pp. 16–22, 128. MR659677

[K2] S. F. Kolyada and A. G. Sivak, *Universal constants for one-parameter families of mappings* (Russian), Oscillation and stability of solutions of functional-differential equations, Akad. Nauk Ukrain. SSR, Inst. Mat., Kiev, 1982, pp. 53–60. MR722803

[K3] S. F. Kolyada and A. G. Sivak, *A class of functional equations and the universal behavior of families of one-dimensional mappings* (Russian), Differential-difference equations and problems of mathematical physics (Russian), Akad. Nauk Ukrain. SSR, Inst. Mat., Kiev, 1984, pp. 33–37, 133. MR884351

[K4] S. F. Kolyada, *Mappings of the interval with a zero Schwarzian* (Russian), Functional-differential equations and their applications (Russian), Akad. Nauk Ukrain. SSR, Inst. Mat., Kiev, 1985, pp. 47–57, iii. MR892977

[K5] S. F. Kolyada, *The measure of quasi-attractors of one-dimensional smooth mappings* (Russian), Akad. Nauk Ukrain. SSR Inst. Mat. Preprint **35** (1986), 22. MR865510

[K6] S. F. Kolyada, *On the dynamics of triangular mappings* (Russian), Akad. Nauk Ukrain. SSR Inst. Mat. Preprint **61** (1986), 28. MR883283

[K7] S. F. Kolyada, *Discrete dynamical systems with a Schwartzian derivative of constant sign* (Russian), Functional-differential equations and their application to nonlinear boundary value problems (Russian), Akad. Nauk Ukrain. SSR, Inst. Mat., Kiev, 1987, pp. 18–23, i. MR950961

[K8] S. F. Kolyada, *One-parameter families of mappings of an interval with a negative Schwartzian derivative which violate the monotonicity of bifurcations* (Russian), Ukrain. Mat. Zh. **41** (1989), no. 2, 258–261, 288, DOI 10.1007/BF01060393; English transl., Ukrainian Math. J. **41** (1989), no. 2, 230–233. MR992831

[K9] S. F. Kolyada, *Triangular mappings of type* 2^∞ *with positive entropy* (Russian), Dynamical systems and turbulence (Russian), Akad. Nauk Ukrain. SSR, Inst. Mat., Kiev, 1989, pp. 76–82, iii. MR1051939

[K10] S. F. Kolyada, *On dynamics of triangular maps of the square* (English, with Russian summary), Akad. Nauk Ukrain. SSR Inst. Mat. Preprint **14** (1991), 36. MR1153035

[K11] S. F. Kolyada and L'. Snoha, *On* ω*-limit sets of triangular maps* (English, with Russian summary), Akad. Nauk Ukrain. SSR Inst. Mat. Preprint **30** (1991), 20. MR1153262

[K12] S. F. Kolyada and A. N. Sharkovsky, *On topological dynamics of triangular maps of the plane*, European Conference on Iteration Theory (Batschuns, 1989), World Sci. Publ., River Edge, NJ, 1991, pp. 177–183. MR1184163

[K13] S. F. Kolyada, *On dynamics of triangular maps of the square*, Ergodic Theory Dynam. Systems **12** (1992), no. 4, 749–768, DOI 10.1017/S0143385700007082. MR1200342

[K14] S. F. Kolyada and L'. Snoha, *On* ω*-limit sets of triangular maps*, Real Anal. Exchange **18** (1992/93), no. 1, 115–130. MR1205504

[K15] L. Alsedà, S. F. Kolyada, and L'. Snoha, *On topological entropy of triangular maps of the square*, Bull. Austral. Math. Soc. **48** (1993), no. 1, 55–67, DOI 10.1017/S000497270001546X. MR1227436

[K16] S. Kolyada and L'. Snoha, *On topological dynamics of sequences of continuous maps*, Proceedings of the Conference "Thirty Years after Sharkovskiĭ's Theorem: New Perspectives" (Murcia, 1994), Internat. J. Bifur. Chaos Appl. Sci. Engrg. **5** (1995), no. 5, 1437–1438, DOI 10.1142/S0218127495001125. MR1361933

[K17] S. Kolyada and L'. Snoha, *On topological dynamics of sequences of continuous maps*, Thirty years after Sharkovskiĭ's theorem: new perspectives (Murcia, 1994), World Sci. Ser. Nonlinear Sci. Ser. B Spec. Theme Issues Proc., vol. 8, World Sci. Publ., River Edge, NJ, 1995, pp. 175–176. Reprint of the paper reviewed in MR1361933 (97b:58076). MR1415894

[K18] S. Kolyada and L'. Snoha, *Topological entropy of nonautonomous dynamical systems*, Random Comput. Dynam. **4** (1996), no. 2-3, 205–233. MR1402417

[K19] S. F. Kolyada and L'. Snoha, *Topological dynamics of triangular maps of the square*, Iteration theory (Batschuns, 1992), World Sci. Publ., River Edge, NJ, 1996, pp. 165–171. MR1442279

[K20] Yu. M. Berezans′kiĭ, O. Yu. Romanenko, V. V. Fedorenko, S. F. Kolyada, M. B. Vērēĭkīna, and A. G. Sīvak, *Oleksandr Mikolaĭovich Sharkovs′kiĭ (on the occasion of his sixtieth birthday)* (Ukrainian), Ukraïn. Mat. Zh. **48** (1996), no. 12, 1602–1603, DOI 10.1007/BF02375369; English transl., Ukrainian Math. J. **48** (1996), no. 12, 1815–1816 (1997). MR1443653

[K21] S. Kolyada and L'. Snoha, *Some aspects of topological transitivity—a survey*, Iteration theory (ECIT 94) (Opava), Grazer Math. Ber., vol. 334, Karl-Franzens-Univ. Graz, Graz, 1997, pp. 3–35. MR1644768

[K22] Ll. Alsedà, S. Kolyada, J. Llibre, and L'. Snoha, *Entropy and periodic points for transitive maps*, Trans. Amer. Math. Soc. **351** (1999), no. 4, 1551–1573, DOI 10.1090/S0002-9947-99-02077-2. MR1451592

[K23] S. Kolyada, M. Misiurewicz, and L'. Snoha, *Topological entropy of nonautonomous piecewise monotone dynamical systems on the interval*, Fund. Math. **160** (1999), no. 2, 161–181. MR1706474

[K24] S. Kolyada, L'. Snoha, and S. Trofimchuk, *Noninvertible minimal maps*, Fund. Math. **168** (2001), no. 2, 141–163, DOI 10.4064/fm168-2-5. MR1852739

[K25] F. Blanchard, E. Glasner, S. Kolyada, and A. Maass, *On Li-Yorke pairs*, J. Reine Angew. Math. **547** (2002), 51–68, DOI 10.1515/crll.2002.053. MR1900136

[K26] H. Bruin, S. Kolyada, and L'. Snoha, *Minimal nonhomogeneous continua*, Colloq. Math. **95** (2003), no. 1, 123–132, DOI 10.4064/cm95-1-10. MR1967559

[K27] E. Akin and S. Kolyada, *Li-Yorke sensitivity*, Nonlinearity **16** (2003), no. 4, 1421–1433, DOI 10.1088/0951-7715/16/4/313. MR1986303

[K28] L. Alsedà, S. Kolyada, J. Llibre, and L'. Snoha, *Axiomatic definition of the topological entropy on the interval*, Aequationes Math. **65** (2003), no. 1-2, 113–132, DOI 10.1007/s000100300008. MR2012405

[K29] S. Kolyada, L'. Snoha, and S. Trofimchuk, *On minimality of nonautonomous dynamical systems* (English, with English and Ukrainian summaries), Nelīnīĭnī Koliv. **7** (2004), no. 1, 86–92, DOI 10.1023/B:NONO.0000041798.79176.94; English transl., Nonlinear Oscil. (N. Y.) **7** (2004), no. 1, 83–89. MR2151023

[K30] S. F. Kolyada, *Li-Yorke sensitivity and other concepts of chaos* (Ukrainian, with English and Ukrainian summaries), Ukraïn. Mat. Zh. **56** (2004), no. 8, 1043–1061, DOI 10.1007/s11253-005-0055-4; English transl., Ukrainian Math. J. **56** (2004), no. 8, 1242–1257. MR2136308

[K31] S. F. Kolyada, *Topological entropy of a dynamical system on the space of one-dimensional mappings* (Ukrainian, with English and Ukrainian summaries), Nelīnīĭnī Koliv. **7** (2004), no. 2, 180–187, DOI 10.1007/s11072-005-0004-z; English transl., Nonlinear Oscil. (N. Y.) **7** (2004), no. 2, 179–186. MR1955085

[K32] J. Auslander, S. Kolyada, and L'. Snoha, *Functional envelope of a dynamical system*, Nonlinearity **20** (2007), no. 9, 2245–2269, DOI 10.1088/0951-7715/20/9/012. MR2351032

[K33] M. Kapranov, S. Kolyada, Y. Manin, P. Moree, and L. Potyagailo, *Preface*, Geometry and dynamics of groups and spaces, Progr. Math., vol. 265, Birkhäuser, Basel, 2008, pp. vii–xiv, DOI 10.1007/978-3-7643-8608-5. MR2394795

[K34] S. Kolyada, L'. Snoha, and S. Trofimchuk, *Proper minimal sets on compact connected 2-manifolds are nowhere dense*, Ergodic Theory Dynam. Systems **28** (2008), no. 3, 863–876, DOI 10.1017/S0143385707000740. MR2422019

[K35] S. Kolyada and L'. Snoha, *Topological transitivity*, Scholarpedia **4(2):5802** (2009), `http://www.scholarpedia.org/article/Topological_transitivity`.

[K36] S. Kolyada and L'. Snoha, *Minimal dynamical systems*, Scholarpedia **4(11):5803** (2009), `http://www.scholarpedia.org/article/Minimal_dynamical_systems`.

[K37] S. Kolyada and M. Matviichuk, *On extensions of transitive maps*, Discrete Contin. Dyn. Syst. **30** (2011), no. 3, 767–777, DOI 10.3934/dcds.2011.30.767. MR2784619

[K38] S. Kolyada and D. Robatian, *On ω-limit sets of triangular induced maps*, Real Anal. Exchange **38** (2012/13), no. 2, 299–315. MR3261879

[K39] S. Kolyada and O. Rybak, *On the Lyapunov numbers*, Colloq. Math. **131** (2013), no. 2, 209–218, DOI 10.4064/cm131-2-4. MR3092451

[K40] E. Bilokopytov and S. F. Kolyada, *Transitive maps on topological spaces*, Ukrainian Math. J. **65** (2014), no. 9, 1293–1318, DOI 10.1007/s11253-014-0860-8. MR3176447

[K41] S. Kolyada, L'. Snoha, and S. Trofimchuk, *Minimal sets of fibre-preserving maps in graph bundles*, Math. Z. **278** (2014), no. 1-2, 575–614, DOI 10.1007/s00209-014-1327-1. MR3267591

[K42] S. Kolyada and J. Semikina, *On topological entropy: when positivity implies +infinity*, Proc. Amer. Math. Soc. **143** (2015), no. 4, 1545–1558, DOI 10.1090/S0002-9939-2014-12282-9. MR3314068

[K43] S. Kolyada, M. Misiurewicz, and L'. Snoha, *Spaces of transitive interval maps*, Ergodic Theory Dynam. Systems **35** (2015), no. 7, 2151–2170, DOI 10.1017/etds.2014.18. MR3394112

[K44] W. Huang, D. Khilko, S. Kolyada, and G. Zhang, *Dynamical compactness and sensitivity*, J. Differential Equations **260** (2016), no. 9, 6800–6827, DOI 10.1016/j.jde.2016.01.011. MR3461085

[K45] S. Kolyada, M. Misiurewicz, and L'. Snoha, *Loops of transitive interval maps*, Dynamics and numbers, Contemp. Math., vol. 669, Amer. Math. Soc., Providence, RI, 2016, pp. 137–154, DOI 10.1090/conm/669/13427. MR3546667

[K46] A. F. Ivanov, S. F. Kolyada, V. L. Maĭstrenko, and et al., *Oleksandr Mikolaĭovich Sharkovs′kiĭ (on his 80th birthday)* (Ukrainian), Ukraïn. Mat. Zh. **69** (2017), no. 2, 257–260. MR3631634

[K47] W. Huang, S. Kolyada, and G. Zhang, *Analogues of Auslander-Yorke theorems for multi-sensitivity*, Ergodic Theory Dynam. Systems **38** (2018), no. 2, 651–665, DOI 10.1017/etds.2016.48. MR3774836

[K48] W. Huang, D. Khilko, S. Kolyada, A. Peris, and G. Zhang, *Finite intersection property and dynamical compactness*, J. Dynam. Differential Equations **30** (2018), no. 3, 1221–1245, DOI 10.1007/s10884-017-9600-8. MR3842147

[K49] S. Kolyada, *A survey of some aspects of dynamical topology: Dynamical compactness and Slovak spaces*, Discrete Contin. Dyn. Syst. Ser. S, 27pp., to appear. DOI 10.3934/dcdss.2020074.

[K50] S. Kolyada, M. Misiurewicz, and L'. Snoha, *Special α-limit sets*, Dynamics: Topology and Numbers, Contemp. Math., vol. 744, Amer. Math. Soc., Providence, RI, 2019, pp. 157–173.

Books by Sergiĭ Kolyada

[K51] A. N. Sharkovskiĭ, S. F. Kolyada, A. G. Sivak, and V. V. Fedorenko, *Dinamika odnomernykh otobrazheniĭ* (Russian), "Naukova Dumka", Kiev, 1989. MR1036027

[K52] A. N. Sharkovsky, S. F. Kolyada, A. G. Sivak, and V. V. Fedorenko, *Dynamics of one-dimensional maps*, Mathematics and its Applications, vol. 407, Kluwer Academic Publishers Group, Dordrecht, 1997. Translated from the 1989 Russian original by Sivak, P. Malyshev and D. Malyshev and revised by the authors. MR1448407

[K53] S. Kolyada, *Topologīchna dinamīka: mīnīmal′nīst′, entropīya ta khaos* (Ukrainian, with Ukrainian summary), Pratsī Īnstitutu Matematiki Natsīonal′noï Akademīï Nauk Ukraïni. Matematika ta ïï Zastosuvannya [Proceedings of Institute of Mathematics of NAS of Ukraine. Mathematics and its Applications], vol. 89, Natsīonal′na Akademīya Nauk Ukraïni, Īnstitut Matematiki, Kiev, 2011. MR3089342

Publications co-edited by Sergiĭ Kolyada

[K54] S. Bezuglyi and S. Kolyada (eds.), *Topics in dynamics and ergodic theory*, London Mathematical Society Lecture Note Series, vol. 310, Cambridge University Press, Cambridge, 2003. MR2052271

[K55] S. Kolyada, Y. Manin, and T. Ward (eds.), *Algebraic and topological dynamics*, Contemporary Mathematics, vol. 385, American Mathematical Society, Providence, RI, 2005. MR2181447

[K56] M. Kapranov, S. Kolyada, Y. I. Manin, P. Moree, and L. Potyagailo (eds.), *Geometry and dynamics of groups and spaces*, Progress in Mathematics, vol. 265, Birkhäuser Verlag, Basel, 2008. In memory of Alexander Reznikov; Including papers from the International Conference held in Bonn, September 22–29, 2006. MR2401786

[K57] S. Kolyada, Y. Manin, M. Möller, P. Moree, and T. Ward (eds.), *Dynamical numbers—interplay between dynamical systems and number theory*, Contemporary Mathematics, vol. 532, American Mathematical Society, Providence, RI, 2010. MR2849686

[K58] S. Kolyada, M. Möller, P. Moree, and T. Ward (eds.), *Dynamics and numbers*, Contemporary Mathematics, vol. 669, American Mathematical Society, Providence, RI, 2016. MR3546660

Other references

[ABCP89] S. J. Agronsky, A. M. Bruckner, J. G. Ceder, and T. L. Pearson, *The structure of ω-limit sets for continuous functions*, Real Anal. Exchange **15** (1989/90), no. 2, 483–510. MR1059418

[BS03] F. Balibrea and L'. Snoha, *Topological entropy of Devaney chaotic maps*, Topology Appl. **133** (2003), no. 3, 225–239, DOI 10.1016/S0166-8641(03)00090-7. MR2000500

[Bow71] R. Bowen, *Entropy for group endomorphisms and homogeneous spaces*, Trans. Amer. Math. Soc. **153** (1971), 401–414, DOI 10.2307/1995565. MR274707

[DM86] R.-A. Dana and L. Montrucchio, *Dynamic complexity in duopoly games*, J. Econom. Theory **40** (1986), no. 1, 40–56, DOI 10.1016/0022-0531(86)90006-2. MR862583

[Dir08] M. Dirbák, *Extensions of dynamical systems without increasing the entropy*, Nonlinearity **21** (2008), no. 11, 2693–2713, DOI 10.1088/0951-7715/21/11/011. MR2448237

[DF54] Y. N. Dowker and F. G. Friedlander, *On limit sets in dynamical systems*, Proc. London Math. Soc. (3) **4** (1954), 168–176, DOI 10.1112/plms/s3-4.1.168. MR0061820

[DST17] T. Downarowicz, L'. Snoha, and D. Tywoniuk, *Minimal spaces with cyclic group of homeomorphisms*, J. Dynam. Differential Equations **29** (2017), no. 1, 243–257, DOI 10.1007/s10884-015-9433-2. MR3612974

[Fan19] X. Fan, J. Li, Y. Yang, and Z. Yang *Subspaces of interval maps related to the topological entropy*, Topol. Methods Nonlinear Anal., to appear. Prepprint published October 7, 2019, DOI 10.12775/TMNA.2019.065.

[Klo79] P. E. Kloeden, *On Sharkovsky's cycle coexistence ordering*, Bull. Austral. Math. Soc. **20** (1979), no. 2, 171–177, DOI 10.1017/S0004972700010819. MR557223

[Mat13] M. Matviichuk, *On the dynamics of subcontinua of a tree*, J. Difference Equ. Appl. **19** (2013), no. 2, 223–233, DOI 10.1080/10236198.2011.634804. MR3021784

[MS80] M. Misiurewicz and W. Szlenk, *Entropy of piecewise monotone mappings*, Studia Math. **67** (1980), no. 1, 45–63, DOI 10.4064/sm-67-1-45-63. MR579440

[Sin78] D. Singer, *Stable orbits and bifurcation of maps of the interval*, SIAM J. Appl. Math. **35** (1978), no. 2, 260–267, DOI 10.1137/0135020. MR0494306

[Šar65] O. M. Šarkovs′kiĭ, *On attracting and attracted sets* (Russian), Dokl. Akad. Nauk SSSR **160** (1965), 1036–1038. MR0188992

[SMR93] A. N. Sharkovsky, Yu. L. Maĭstrenko, and E. Yu. Romanenko, *Difference equations and their applications*, Mathematics and its Applications, vol. 250, Kluwer Academic Publishers Group, Dordrecht, 1993. Translated from the 1986 Russian original by D. V. Malyshev, P. V. Malyshev and Y. M. Pestryakov. MR1244579

[Šte16] M. Štefánková, *The Sharkovsky program of classification of triangular maps—a survey*, Topology Proc. **48** (2016), 135–150. MR3352336

[Zdu84] A. Zdunik, *Entropy of transformations of the unit interval*, Fund. Math. **124** (1984), no. 3, 235–241, DOI 10.4064/fm-124-3-235-241. MR774514

Department of Mathematics, Faculty of Natural Sciences, Matej Bel University, Tajovského 40, 974 01 Banská Bystrica, Slovakia

Email address: Lubomir.Snoha@umb.sk

Contemporary Mathematics
Volume **744**, 2020
https://doi.org/10.1090/conm/744/14984

Recollections about Sergiĭ Kolyada

Irina Kolyada

I'm the eldest daughter of a great man, a remarkable mathematician and best father in the world – Sergiĭ Kolyada. Many people know the scientific achievements of my father the Professor of Mathematics, but I would like to say something about his life.

I'm not a mathematician, but my father told me that mathematics is an incredible science, very hard and very beautiful. His love for mathematics started early in his childhood. He was born in a very small village and studied at a village school, where the teacher of mathematics saw in a young boy a talented, diligent and persistent mathematician. Due to this, at the age of 15, that boy from the village went to a mathematical school in Kyiv. I could talk for a long time about how he won mathematical Olympiads, and how with the support of his teacher he was able to get to Kyiv, but if we could ask Sergiĭ "How did you get to mathematics?", his answer would be unequivocal: "I loved mathematics", and true love does miracles.

Dad told us that his school years were difficult and, at the same time, very interesting. Teachers of this school were preparing real mathematicians, not just for the program, nor to win Olympiads, but for the principle that "the tasks must always be pursued by mathematicians, a professional mathematician is one whose mathematical problems chase days, months, years! Otherwise, you are not a mathematician."

Later, he studied at the Taras Shevchenko National University, in the Faculty of Mechanics and Mathematics. There Sergiĭ Kolyada met his future wife and the only love of all his life, Mariya. I should mention the interesting arithmetical fact that my mother's birthday is March 14, a date shared with π.

As a student, Sergiĭ regularly participated in the seminars of the Institute of Mathematics of the National Academy of Sciences of Ukraine, despite the fact that he only started to work there seven years later.

Personally, I have incredible memories from my childhood. Even now, when I go to the Institute, I see a little girl, who knew every step of this building, every crack of the floor, every employee. I adored it when my dad took me to his work, as for me it had a special smell of mathematics. That was the period when papers were prepared on typewriter, with copies through carbon paper or written in notebooks by hand, and formulas were recorded with chalk on the board. I especially liked to make corrections in my father's papers "$-$" replaced with "$+$", "x" with "y". Of course dad noticed, and joked that I forced him to be more attentive.

Sergiĭ's family (photo with permission of Irina Kolyada)

I remember the first time I went to an international conference, which was headed by my father. Given the atmosphere that mathematicians inhabited, I was proud to be a daughter of a professor of mathematics.

Especially memorable were our visits to France, Germany, Spain,... I appreciated the respect the wider mathematical community showed to Professor Kolyada! He prepared with great care and inspiration for every conference, piece of work and speech.

Despite the fact that during the last 15 years my father worked outside Ukraine, he was a real patriot who devoted himself to the development of mathematics there. Doing science in Ukraine is not easy. Dad, with like-minded collaborators, created the Kyiv Mathematical Society, and for many years was the President of the society, whose main task was to popularize mathematics as a science. I particularly want to emphasize the special contribution he made to the support of young mathematicians of Ukraine.

In 2006 Sergiĭ began to teach at the Taras Shevchenko National University, where he had studied. Whatever the weather in Kyiv, whatever the circumstances, on Monday, after a two-hour journey, dad reached the audience to conduct seminars for students for a penny. I asked him "Why do you need it?", and the answer was very simple: "If at least a few of my students continue my work, I will be happy to have left a mark in their life and the life of the future mathematics of Ukraine." And it happened. Over the past eight years, my father had students to whom he just gave "problems" at University, not expecting any solutions. The result was that with almost all of his students he published scientific work. All of them are now continuing his work, studying at postgraduate courses in Bonn, Paris and other European cities.

In addition, he founded and became co-chair of the Competition of the Taras Shevchenko Scientific Society in America and the Ukraine–USA Foundation for

Young Mathematicians in Ukraine. He also initiated the creation of grants named by M. Ostrogradsky for the ten best student mathematicians of Universities in Kyiv. It is a great contribution to the future of mathematics in Ukraine.

Sport was dad's real hobby. In his student years, he played for the university basketball team. He supported his brother, the well-known Ukrainian hockey player Mykola Kolyada. And of course, every day for many years he supported my sister Natalia, now a world champion in sport ballroom dancing.

Over the last 15 years he played football. In almost every place, where he lived or worked, he joined a football team. In Kyiv it was the Football Club "Fortuna", and twice a week they played their favorite game no matter what the weather was. It was in one of those games that he spent the last minutes of his life.

Whether he was thinking of mathematics or playing football, the family was always at the front of his mind. He was proud of the achievements of his daughters and grandchildren, and always loved and supported his wife. Grandchildren admired him and always wanted to be like him. And for me and my sister he was definitely the best dad in the world. Every day, wherever he was, we all received a message – "Good morning and have a good day!". The same message was received on May 16 of 2018. The last one.

Alexander (Sasha) Blokh

I have known Sergiĭ since around 1973, when we both participated in the all-Ukrainian and all-Soviet mathematical Olympiads; he was a member of the Poltava team, and I was a member of the Kharkiv team. During the competition, it just so happened that we stood next to each other in line at a self-service restaurant. We shook hands, and step by step started talking about a variety of topics, from sports to mathematics. We became friends and have stayed friends ever since.

We kept in touch through letters after we entered colleges (he in Kyiv, I in Kharkiv), and began discussing more advanced mathematics; evidently, we had similar taste in it as both of us got interested in dynamical systems, in particular, in one-dimensional dynamics (maybe because we were both impressed by the Sharkovskiĭ theorem), so our professional contacts continued at a new level.

However there was a lot of human warmth in this interaction too. I remember vividly how Sergiĭ stayed with us in Moscow for a couple of days when our first son Ilya was 4 years old. Sergiĭ entertained Ilya by telling him and all of us funny stories and riddles. For some time afterwards Ilya would ask me whether this funny "dyadya" (which literally means "uncle" but in this context simply means "man") was coming back.

I moved to the US in 1991, still we would often meet at conferences or meetings, including various activities at Max-Planck Institut für Mathematik in Bonn where Sergiĭ would invite me. I remember visiting him in Kyiv some time in the 1990-s; we had two unforgettable nights just walking in this great city, listening to street musicians, and talking.

You may not know this, but Poltava is a nice city of 300,000 in the heart of Ukraine. It gave the world a number famous people such as philosopher Hryhoriy Skovoroda or writer Nikolay Gogol; a well-known mathematician named Ostrogradsky is from Poltava too. To me, though, the most important fact about Poltava is that modern Ukrainian literature was born in Poltava when Ivan Kotlyarevsky published his mock-heroic poem *Eneyida*. The soft and beautiful Poltava dialect gave rise to the modern literary Ukrainian language, and this was the language I loved to hear from Sergiĭ, who spoke it very eloquently. I will always remember him just like that, as a soft-spoken, always smiling, kind and helpful person, and a great friend.

Ľubomír Snoha

Sergiĭ was my most important co-author – we wrote some 20 joint papers. We were also close friends. When one of us needed help, either in work or in a private matter, or simply needed a word of support, he could always turn to the other one who, for sure, did his best to help him, putting aside all other work and duties.

We often used to meet at conferences and joint research stays in Banská Bystrica, in Kiev and in research institutions abroad. Our families also knew each other well. I lived with his family when in Kiev, and he with us in Banská Bystrica.

The first time we met was from May 29 to June 2, 1989 during the 'Summer School on Dynamical Systems – Bratislava, Kiev, Warsaw' in Modra-Piesky, Slovakia, and it was friendship at first sight. Apart from some short visits I made to Banská Bystrica, Sergiĭ had been there earlier. Indeed, in July 1978, as a university student, he took part in the so-called 'International Summer Brigadier Camp' in Dubová near Banská Bystrica, while I started working at the Faculty of Education in Banská Bystrica only a few days later, since August 1978. The summer camp, part of the so-called 'student summer activity' organized by the Socialist Youth Union, brought students from different Soviet block countries to work in factories in the district of Banská Bystrica. Sergiĭ worked at the construction site at the Petrochema Dubová refinery, and we often laughed at memories of his experiences. Ukrainian students, unlike students from some other countries who were a bit lazy, worked really hard and therefore made very good money. After the brigade they went on a sightseeing trip to Prague, then finally back to Kiev.

Later in June 1990 we met at the '3rd Czechoslovak Summer School on Dynamical Systems' in Dubník (near Stará Turá).

At this point Sergiĭ and I were 'only' friends and were not cooperating in mathematics. We became co-authors through Jaroslav Smítal. I was a former student of Jaroslav and I regularly visited his seminar on dynamical systems at my alma mater, Comenius University in Bratislava. In 1990 he offered me a visit to Kiev under a scientific agreement between Comenius University and the Mathematical Institute of the Ukrainian Academy of Sciences. At that time Sergiĭ was preparing the English version of his results on triangular maps. We discussed the topic and I also helped him with the preparation of the paper and in particular with English. (At that time his knowledge of English was close to zero; it is remarkable that later he was able to improve so much.) His paper then appeared in Ergodic Theory and Dynamical Systems in 1992.

During my stay we started to collaborate. Since he was preparing his paper on triangular maps and before my visit to Kiev I read an article on omega-limit sets, I asked him whether he could characterize omega-limit sets of triangular maps, or at least omega-limit sets lying in one fibre. He did not, so we started to think about the problem. Of course, we found time also for other activities. Sergiĭ, as a football fan, invited me to a stadium in Kiev, to watch the footbal match between Dynamo Kiev and CSKA Moscow. Dynamo won 4-1.

Then Sergiĭ first visited me. He stayed in a student dormitory. When I suggested we could go and see my family, he was concerned that my wife would not welcome an unexpected visitor. I told him something like "А ты думаеш што мы не нормальные люди?" ("Do you think that we are not normal people?"), and he recalled this story many years later.

At that time we did not have access to email communication. From time to time we called each other, but it was very expensive and one had to wait a long time at the post office for a connection. So, we used to send letters to each other, but it was too slow and the conference in Kiev where we wanted to present the not yet completed result, was already close. It was necessary to use non-traditional methods of communication. Once I traveled to Bratislava airport, and approached a young man in the queue to check-in for a flight to Kiev. He agreed to take a letter to Sergiĭ, after I showed him that it was just mathematics and nothing dangerous. I showed him a picture of Sergiĭ, saying that he would wait at the airport in Kiev. Sergiĭ heroically typed our work in ChiWriter in the form of a preprint of the institute. During the Soviet-Spanish-Czechoslovakian Symposium 'Dynamical Systems and their Applications' in Kiev, June 28 - July 4, 1991, the preprint was already available and I gave a talk on our first joint work 'On omega-limit sets of triangular maps', in part made possible by a kind young traveller.

Our first joint research stay outside our countries was a month stay in Barcelona at the very beginning of 1993. We were invited by Lluis Alsedà and Jaume Llibre to the Centre de Recerca Matemàtica. We all four investigated mainly the entropy bounds problem in various spaces (what is the infimum of the topological entropy for transitive maps in a given space?). Lluis and Jaume obtained early results, while we were trying to solve the problem for triangular maps. In the middle of our stay we still did not have any contribution, but one night, at about two o'clock, sitting at the table and working hard in our flat in San Cugat, we succeeded. When looking back at our careers, this stay was very important and we vere grateful to our both hosts. We used to buy some food in a shop close to our flat, and we were disappointed with the fact that, as it seemed to us, in Spain there was no bread. Instead, we were buying 'pan de molde' which was too soft for us. By chance, in the second half of our stay, we realized that we had used a flawed inductive argument 'if there is no bread in our shop, then there is no bread in any other shop'. In fact, in a nearby bakery there was a good choice of various kinds of bread.

In 1996 we spent a month doing 'Research in Pairs' in Oberwolfach, and Sergiĭ travelled some 57 hours on buses and trains to get there. At that time we could not afford flights. This was quite usual for us in those years, and I once took three days to get to a conference in Lisbon.

Sergiĭ was a real friend who always was thinking about the others and followed the rule 'Guest in the house, God in the house.' In the 1990's I had health problems with asthma and allergy. Sergiĭ of course knew about that and he also knew that because of that we had removed all the carpets in my flat in Banská Bystrica. In those years I visited him and stayed in the room of his daughters while the girls lived with their parents in the rest of the flat. To my surprise they also had no carpets, and I only recently learned that they removed them just for the period of my stay! This consideration was typical of Sergiĭ, and on a later visit in May 1994 he arranged for us to work together on the coast in Katsiveli in Crimea for the benefit of my health. Thanks to Sergiĭ, I could breath the sea air. We rented a room there and did mathematics on the beach, with stones put on the sheets of papers, to stop the wind stealing them.

The journey to Crimea was interesting. Since we both were fans of the satirical novels 'The Twelve Chairs' and 'The Little Golden Calf' by Ilf and Petrov, we knew some parts of them by heart. In particular, Balaganov, one of the characters of the latter novel, knew, apart from Moscow, out of all the 'seats of world culture' only Kiev, Melitopol and Zhmerinka. Of course, when our train stopped at Melitopol, we both got off the train to stand for at least a moment on the ground of this 'seat of world culture'. Later Sergiĭ and I also always stopped for a while at the monument to Panikovsky, another character from 'The Little Golden Calf', when I was in Kiev.

Once we visited my mother in Lučenec in south of Slovakia, because Sergiĭ had a better bus connection to Ukraine from there than from Banská Bystrica. Having arrived at Lučenec, I helped him with his luggage and somehow I forgot my bag in the bus. This was a catastrophe, because our notes with what we were doing last two weeks were in the bag and so they were lost. We both went back to the bus station, but the bus was not there and it was getting dark. From some other drivers I got the information that the bus was in a depot somewhere behind Lučenec. By taxi we were able to find the place. There were buses and two dogs behind the gate, and nothing like a doorbell. Fortunately the dogs were barking so loudly that finally a night watchman appeared and already from afar he cried: 'The bag is here!' So we were lucky and our mathematics was saved.

Sergiĭ is waving at us
(photo with permission of Michał Misiurewicz)

Sergiĭ was a practical man, while I am not. In particular, when we were together somewhere (say in Barcelona, Marseille or Bonn) and we were preparing food in the kitchen, he was the cook, and I the assistant. My mother and my wife always asked me: "Will Sergiĭ also be there?" when I travelled. They knew that Sergiĭ would take good care of their son and husband. Sergiĭ was very well liked by my family, and after his death my mother asked her priest to serve a mass for him.

We generally spoke Russian to each other. However, my mother and Sergiĭ spoke Slovak and Ukrainian and claimed to understand each other. Sergiĭ sometimes laughed at mistakes in my Russian, and I used to repeat some mistakes on purpose, as they somehow belonged to us. However, if anybody asked him about my Russian, he always claimed that it was very good.

I miss my great friend Sergiĭ very much and I will not forget about him till the end of my days.

Contemporary Mathematics
Volume **744**, 2020
https://doi.org/10.1090/conm/744/14977

Sergiy and the MPIM

Pieter Moree

Sergiy was a regular visitor of the MPIM, almost part of the furniture. Between 2000 and 2017 he was in Bonn almost every year, each time for a period of several months. In this way he spent over the years a total of about 5 years at MPIM!

I got to know Sergiy in 2004 (the year I started to work at MPIM), when he organized an activity. Our contact became much more intensive when we were co-organizing the 2006 memorial conference for Alexander Reznikov [**6**] (a talented Kiev born mathematician who faced many difficulties in his life, and who unfortunately, like Sergiy, passed away very early). Co-organizing that conference with Sergiy was dear to me, as it was a small way to pay tribute to a coauthor (see [**9**]) I admired greatly.

F. Hirzebruch with Ukrainian mathematicians A. Kosyak,
S. Kolyada, M. Viazovska and V. Lyubashenko after becoming
a Honary Member of the Kiev Mathematical Society
(with permission of the Max Planck Institute for Mathematics)

After my first co-organizational experience with Sergiy, several more were to follow, some of these quite time intensive. It is therefore not surprising that my contact with Sergiy was usually focused on pushing further the organization of the activities, rather than other matters. A particularly helpful feature of working with Sergiy was his uncanny ability to respond to email requests almost immediately. Of course, when he was physically at MPIM and the issue was not absolutely minor, I would go to his office — always the same one, his favorite — to discuss things in person. Often enough I would run into him at the coffee machine, where he would be taking his usual coffee with lime.

For Sergiy and me some of the most relaxing organizational moments always came with the standard conference boat trip on the Rhine, to various destinations. As far as I remember we always had wonderful summer weather.

Over the years it was always clear to me that his family was very dear to him, and a source of much pride. Sergiy was also a very keen soccer player. Indeed, he organized soccer matches on the Hofgarten, a pleasant grass field not far away from MPIM. A few times I joined in, and team was jokingly called FC Hofgarten. The young ones would run and run, but Sergiy would position himself strategically, and in this way was often able to score. His organizational and personal skills showed themselves in soccer as well as in mathematics. In later years he also played indoor soccer, with a largely German group of people not particularly having a scientific or mathematical background.

Somehow outside the institute we rarely met in the course of joint lunches or dinners. It thus came as a surprise — a very memorable one for me — to be invited with a small group to the restaurant *La Cigale* on Friedrichstraße for Sergiy's 50th birthday celebration. Each guest in turn toasted his future and spoke some personal words about Sergiy. He was quite amazed about how fast time flies, expressed his deep gratitude to his friends, and was quite moved by the toasts. This is a lunch I will never forget. There were also some joint lunches with Prof. F. Hirzebruch [**1**, **11**], during which the deep respect of Sergiy for Prof. Hirzebruch was clear. Prof. Hirzebruch played a vital role in reconnecting German mathematicians after the second world war with non-German ones, and he had a huge influence in making Bonn the foremost town in Germany for pure mathematics. A research stay in the early 1950s at the Institute for Advanced Study (IAS) in Princeton inspired Prof. Hirzebruch to found a similar institute in Germany. Eventually he was successful, and the MPIM started in 1982. In 1999 it moved to its current location, right in the center of the city of Bonn. Prof. Hirzebruch always maintained good connections with mathematicians in Eastern European countries, even during the period when this was not at all straightforward. This was amply demonstrated when he managed to get a mathematical star from Moscow, Prof. Yu. I. Manin, to join the MPIM as a Director. As a retired director, Prof. Manin is still a regular presence in the institute and keeps his weekly seminar running.

At a special 4p.m. institute tea on the occasion of the 80th birthday of Prof. Hirzebruch (17 October, 2007), Sergiy handed him a document stating that Prof. Hirzebruch had been made an Honorary Member of the Kiev Mathematical Society.

Sergiy played a crucial role in suggesting highly gifted students from Kiev to apply for a PhD position at the MPIM. These were: Anton Mellit, Danylo Radchenko, Julia Semikina, Maryna Viazovska and Masha Vlasenko. All of them did their PhD with Don Zagier, except Julia who is doing her PhD with Wolfgang Lück. Among

Sergiy and Ľubomír Snoha working at MPIM
(with permission of the Max Planck Institute for Mathematics)

these talented mathematicians, perhaps Maryna Viazovska has found the biggest international acclaim to date. She achieved a breakthrough on optimal sphere packing in dimension 8 that rightly attracted much attention, both mathematical and from the wider media. Later, involving amongst others Danylo Radchenko (who at the time was studying for his PhD at MPIM with Don Zagier and is a current post-doc at MPI) the work was generalized to dimension 24 [**3**]. These two papers are mentioned in the award announcement of the 2017 Sastra Ramanujan prize which Viazovska received. An issue of the American Mathematical Society Notices had her picture as cover, and an article on her work [**4**]. In December 2017 she received a 'New Horizons in Mathematics' prize, which is a subcategory of the Breakthrough in Mathematics Prize. She is honored in the citation for a 'remarkable application of the theory of modular forms to the sphere packing problem in special dimensions.' She also won the European Prize in Combinatorics 2017. In the *laudatio* for that prize, her work on spherical designs [**2**] (finished at MPI) was also mentioned. This work was started in Kiev, but completed whilst she was at MPI.

I already mentioned that Danylo Radchenko and Julia Semikina are both currently employed in Bonn. Masha Vlasenko is a research fellow at the Institute of Mathematics of the Polish Academy of Sciences (IMPAN) in Warsaw, and Anton Mellit an assistant professor at the University of Vienna.

Apart from his own mathematical work, Sergiy's legacy to mathematics is great, and many profited from his hard work and organizational skills, his instinctive kindness, and his wisdom as a person.

0.1. Activities and conferences co-organized at MPIM. This is the list of the activities co-organized by Sergiy, for all of which he was the **originator** of the idea to organize them.

- **Algebraic and topological dynamics**, 2004, Activity: May-July, Conference: July 5-9, 2004 [**5**].
- **Geometry and Dynamics of Groups and Spaces** (in memory of Alexander Reznikov), Sept. 22-29, 2006 [**6**].
- **Dynamical Numbers: Interplay between Dynamical Systems and Number Theory**, Activity: May-July 2009, Conference: July 20-24, 2009 [**7**].
- **Dynamics and numbers**, Activity: July 2014, Conference: July 21-25, 2014 [**8**].
- **Dynamics: Topology and Numbers**, July 2-6, 2018.

All these activities and conferences were sponsored by the MPIM. The 2004 activity was in addition supported by the European Science Foundation. The 2009 activity had as co-sponsors the Clay Institute and the Hausdorff Centre of Mathematics.

It is not common for conferences at the MPIM to be followed up by conference proceedings, and the fact that all of the above led to conference proceedings is in large part due to the great energy, enthusiasm and very substantial efforts of Sergiy. The present editors did not have to deliberate long about whether to bring out a conference proceedings, as producing one we see as a clear action in the spirit of Sergiy and a suitable acknowledgment of his enormous editorial efforts.

Sergiy also helped to conceive and organize the conference in 2018, but since it was to be for his 60th birthday, he did not want to be an official organizer. It is a cruel twist of fate that Sergiy passed away less than two months before this conference.

The conference [**6**] was in memory of Alexander Reznikov (1960-2003). The remaining ones belong to the same series so to say, with the note that due to intensified collaboration with the Hausdorff Institute of Mathematics that is focused on trimester programs, people wanting to apply for an activity at MPIM are urged to apply with HIM. It is for this reason that in 2018 only a conference was organized and not an activity and a conference.

The activities typically had around 100 MPIM external participants, and as a whole form a remarkable account of a rich field of mathematics and a fitting memorial to a remarkable mathematician.

References

[1] *Friedrich Hirzebruch (1927–2012)*, Notices Amer. Math. Soc. **61** (2014), no. 7, 706–727, DOI 10.1090/noti1145. Michael Atiyah and Don Zagier, coordinating editors. MR3221122

[2] Andriy Bondarenko, Danylo Radchenko, and Maryna Viazovska, *Optimal asymptotic bounds for spherical designs*, Ann. of Math. (2) **178** (2013), no. 2, 443–452, DOI 10.4007/annals.2013.178.2.2. MR3071504

[3] Henry Cohn, Abhinav Kumar, Stephen D. Miller, Danylo Radchenko, and Maryna Viazovska, *The sphere packing problem in dimension 24*, Ann. of Math. (2) **185** (2017), no. 3, 1017–1033, DOI 10.4007/annals.2017.185.3.8. MR3664817

[4] Henry Cohn, *A conceptual breakthrough in sphere packing*, Notices Amer. Math. Soc. **64** (2017), no. 2, 102–115, DOI 10.1090/noti1474. MR3587715

[5] Sergiy Kolyada, Yuri Manin, and Thomas Ward (eds.), *Algebraic and topological dynamics*, Contemporary Mathematics, vol. 385, American Mathematical Society, Providence, RI, 2005. MR2181447

[6] Alexander Reznikov, *Analytic topology of groups, actions, strings and varieties*, Geometry and dynamics of groups and spaces, Progr. Math., vol. 265, Birkhäuser, Basel, 2008, pp. 3–93, DOI 10.1007/978-3-7643-8608-5_1. MR2402403

[7] Sergiy Kolyada, Yuri Manin, Martin Möller, Pieter Moree, and Thomas Ward (eds.), *Dynamical numbers—interplay between dynamical systems and number theory*, Contemporary Mathematics, vol. 532, American Mathematical Society, Providence, RI, 2010. MR2849686

[8] Sergiĭ Kolyada, Martin Möller, Pieter Moree, and Thomas Ward (eds.), *Dynamics and numbers*, Contemporary Mathematics, vol. 669, American Mathematical Society, Providence, RI, 2016. MR3546660

[9] Alexander Reznikov and Pieter Moree, *Three-manifold subgroup growth, homology of coverings and simplicial volume*, Asian J. Math. **1** (1997), no. 4, 764–768, DOI 10.4310/AJM.1997.v1.n4.a6. MR1621574

[10] Maryna S. Viazovska, *The sphere packing problem in dimension 8*, Ann. of Math. (2) **185** (2017), no. 3, 991–1015, DOI 10.4007/annals.2017.185.3.7. MR3664816

[11] Don Zagier, *Life and work of Friedrich Hirzebruch*, Jahresber. Dtsch. Math.-Ver. **117** (2015), no. 2, 93–132, DOI 10.1365/s13291-015-0114-1. MR3352625

Max-Planck-Institut für Mathematik, Vivatsgasse 7, D-53111 Bonn, Deutschland

Contemporary Mathematics
Volume **744**, 2020
https://doi.org/10.1090/conm/744/14978

Homotopy types and geometries below Spec($\mathbb{Z}$)

Yuri I. Manin and Matilde Marcolli

To the memory of Sergiy Kolyada

Abstract. After the first heuristic ideas about “the field of one element” $\mathbb{F}_1$ and “geometry in characteristics 1” (J. Tits, C. Deninger, M. Kapranov, A. Smirnov et al.), were developed several general approaches to the construction of “geometries below Spec $\mathbb{Z}$”. Homotopy theory and the “the brave new algebra” were taking more and more important places in these developments, systematically explored by B. Toën and M. Vaquié, among others.

This article contains a brief survey and some new results on *counting problems* in this context, including various approaches to zeta–functions and generalised scissors congruences.

We introduce a notion of $\mathbb{F}_1$ structures based on quasi-unipotent endomorphisms on homology. We also consider $\mathbb{F}_1$ structures based on the integral Bost–Connes algebra and its endomorphisms. In both cases we consider lifts of these structures, via an equivariant Euler charactetristic, to the level of Grothendieck rings and further lifts, via the formalism of assembler categories, to homotopy theoretic spectra.

1. Brief summary and plan of exposition

1.1. Geometries below Spec $\mathbb{Z}$: a general categorical framework. Following [**ToVa09**], Sec. 2.2 – 2.5, we start with a symmetric monoidal category with unit $(C, \otimes, \mathbf{1})$. The category of commutative associative unital monoids $Comm\,(C)$ will play the role of commutative rings; accordingly, the opposite category $Aff_C := Comm\,(C)^{op}$ will be an analogue of the category of *affine schemes.*

In order to be able to define more general schemes, objects of a category Sch_C, we must introduce upon Aff_C *a Grothendieck topology* by giving a collection of sieves (covering families) defined for each object of Aff_C.

It is shown in [**ToVa09**] that if $(C, \otimes, \mathbf{1})$ is complete, cocomplete and closed, then there are several natural topologies upon Aff_C, whose names encode the similarities between them and respective topologies on the category of usual affine schemes (spectra of commutative rings), in particular the *Zariski topology.* Starting

2010 *Mathematics Subject Classification.* Primary 11M26, 14G40, 14G15, 37P35; Secondary 14C15, 14A22, 55P43, 82B10.

Key words and phrases. Weil numbers, quasi-unipotent endomorphisms, Bost–Connes systems, zeta functions, motivic integration, Grothendieck ring of varieties, geometries below Spec $\mathbb{Z}$, assemblers and spectra.

The second author is partially supported by NSF grant DMS-1707882, NSERC Discovery Grant RGPIN-2018-04937 and NSERC Accelerator Supplement grant RGPAS-2018-522593.

with the monoidal symmetric category of abelian groups ($\mathbb{Z}$–Mod, $\otimes_{\mathbb{Z}}, \mathbb{Z}$) one comes to the usual category of schemes $Sch_{\mathbb{Z}}$.

The version of $\mathbb{F}_1$–schemes $Sch_{\mathbb{F}_1}$ suggested in [**ToVa09**] is embodied in the final stretch of the similar path starting with $(Sets, \times, \{*\})$.

Finally, some pairs of categories (Sch_C, Sch_D) can be related by two functors going in opposite directions and satisfying certain adjointness properties. Intuitively, one of them describes appropriate *descent data* upon certain objects X_C of Sch_C necessary and sufficient to define an object X_D lying *under* X_C. One of the most remarkable examples of such descent data from $Sch_{\mathbb{Z}}$ to $Sch_{\mathbb{F}_1}$ was developed by J. Borger, cf. [**Bo11a**], [**Bo11b**]: roughly speaking, it consists in lifting all Frobenius morphisms to the respective $\mathbb{Z}$–scheme. In a weakened form, when only a subset of Frobenius morphisms is lifted, it leads to geometries below $\mathrm{Spec}\,\mathbb{Z}$ but not necessarily over $\mathbb{F}_1$.

1.2. Schemes in the brave new algebra. In a very broad sense, the invasion of mathematics by homotopy theory in general started with a radical enrichment of the Cantorian intuition about what are natural numbers $\mathbb{N}$: they are cardinalities of not just arbitrary finite sets, but rather of sets of connected components π_0 of topological spaces.

The multiplication and addition in $\mathbb{N}$ have then natural lifts to the world of stable homotopy theory, the ring of integers being enriched by passing to the sphere spectrum, where it becomes the initial object, in the same way as $\mathbb{Z}$ itself is an initial object in the category of commutative rings, etc. More details and references are given in Section 4 of this article.

Moreover, "counting functions", such as numbers of $\mathbb{F}_q$–points of a scheme reduced modulo a prime p, with $q = p^a$, can be generalized to the world of scissors congruences where they become the basis for the study of zeta functions.

1.3. The structure of the article. (A). Section 2 is dedicated to a categorification in homotopy theory of a class of schemes in characteristics 1 admitting the following intuitive description: *1–Frobenius morphisms acting upon their cohomology have eigenvalues that are roots of unity.*

Here are some more details. The main arithmetic invariant of an algebraic manifold V defined over a finite field $\mathbb{F}_q$, is its zeta–function $Z(V, s)$ counting the numbers of its points $\mathrm{card}\, V(\mathbb{F}_{q^n})$ over finite extensions of $\mathbb{F}_q$.

Assuming for simplicity that V is irreducible and smooth, as a consequence of the Weil conjectures proved by Deligne, we can identify $Z(V, s)$ with a rational function of q^{-s} which is an alternating product of polynomials whose roots are characteristic numbers of the Frobenius endomorphism Fr_q of V acting upon étale cohomology $H^*_{et}(V)$ of V.

In various versions of $\mathbb{F}_1$–geometry, the structure consisting of cohomology with the action of a Frobenius upon it is conspicuously missing, although it is clearly lurking behind the scene (see e.g. a recent survey and study [**LeBr17**]).

Our homotopical approach here develops the analogy between Frobenius maps and Morse–Smale diffeomorphisms.

In the remaining parts of the article, we focus upon another "counting formalism" and its well developed homotopical environment.

Namely, on the $\mathbb{F}_q$– schemes of finite type, non–necessarily smooth and proper ones, the function $c : V \mapsto \operatorname{card} V(\mathbb{F}_q)$ satisfies the following "scissors identity": if $X \subset V$ is a Zariski closed subscheme, $Y : V \setminus X$, then

$$c(V) = c(X) + c(Y)$$

and also $c(V_1 \times V_2) = c(V_1)c(V_2)$. In other words, c becomes a ring homomorphism of the Grothendieck ring K_0 of the category of $\mathbb{F}_q$–schemes.

(B). In Section 3 we consider a lift to the equivariant Grothendieck ring $K_0^{\hat{\mathbb{Z}}}(\mathcal{V})$ of the integral Bost–Connes system, described in [**CCM09**] in the context of $\mathbb{F}_1$-geometry. We obtain endomorphisms σ_n and additive maps $\tilde{\rho}_n$ that act on $K_0^{\hat{\mathbb{Z}}}(\mathcal{V})$ and map to the corresponding maps of the integral Bost–Connes system through the equivariant Euler characteristic. We obtain in this way a noncommutative enrichment $\mathbf{K}_0^{\hat{\mathbb{Z}}}(\mathcal{V})$ of the Grothendieck ring and an Euler characteristic, which is a ring homomorphism to the integral Bost–Connes algebra. After passing to $\mathbb{Q}$–coefficients, both algebras become semigroup crossed products by the multiplicative semigroup of positive integers,

$$\mathbf{K}_0^{\hat{\mathbb{Z}}}(\mathcal{V}) \otimes_{\mathbb{Z}} \mathbb{Q} = K_0^{\hat{\mathbb{Z}}}(\mathcal{V})_{\mathbb{Q}} \rtimes \mathbb{N},$$

with $K_0^{\hat{\mathbb{Z}}}(\mathcal{V})_{\mathbb{Q}} = K_0^{\hat{\mathbb{Z}}}(\mathcal{V}) \otimes_{\mathbb{Z}} \mathbb{Q}$ and target of the Euler characteristic the rational Bost–Connes algebra $\mathbb{Q}[\mathbb{Q}/\mathbb{Z}] \rtimes \mathbb{N}$.

(C). In Section 4 we revisit the construction of the previous section, by further lifting it from the level of the Grothendieck ring to the level of spectra. We use the approach based on assemblers, developed in [**Za17a**], [**Za17b**].

We recall briefly the general formalism of [**Za17a**] and present a small modification of the construction of [**Za17c**] of the assembler and spectrum associated to the Grothendieck ring of varieties, which will be useful in the following, namely the case of the equivariant Grothendieck ring $K_0^{\hat{\mathbb{Z}}}(\mathcal{V})$ considered in the previous section.

We then prove that the lift of the integral Bost–Connes algebra to the level of the Grothendieck ring described in the previous section can further be lifted to the level of spectra.

(D). In Section 5 we discuss the construction of quantum statistical mechanical expectation values on the Grothendieck ring $K_0^{\hat{\mathbb{Z}}}(\mathcal{V})$ based on motivic measures and the expectation values of the Bost–Connes system.

(E). Finally, in Section 6 we revisit the construction discussed in the previous section in a setting where, instead of considering varieties with a good $\hat{\mathbb{Z}}$-action, we consider a "dynamical" model of $\mathbb{F}_1$-structure based on the existence of an endomorphism $f : X \to X$ that induces a quasi–unipotent map f_* in homology. Our purpose here is to show a compatibility between this proposal about $\mathbb{F}_1$-structures and the idea of [**CCM09**] of $\mathbb{F}_1$-geometry encoded in the structure of the integral Bost–Connes algebra, through its relation to cyclotomic fields.

This also returns us to the framework of Section 2, thus closing the circle.

1.4. General context and goals. As we have already mentioned, behind the many different manifestations of $\mathbb{F}_1$-geometry there is a general unifying idea that can be roughly stated as the principle that roots of unity and cyclotomic points should provide a good notion of points over $\mathbb{F}_1$ and extensions, as well as associated counting functions. This idea was made explicit in the unpublished notes [**KapSmi**] and developed in various directions in [**Ma95**] and [**Sou04**] (understanding the nature of the archimedean Euler factors) and later in [**CC10**] and [**Ma10**]. This cyclotomic viewpoint of $\mathbb{F}_1$-geometry was reformulated in [**CCM09**] in terms of the Bost–Connes quantum statistical mechanical system, where the endomorphisms of the Bost–Connes algebra and their action on roots of unity are seen as a replacement for the classical Frobenius action in the case of points over finite fields. It was further shown in [**Mar09**] that this Bost–Connes formulation of $\mathbb{F}_1$-geometry is compatible with the approach to $\mathbb{F}_1$-geometry via λ-rings developed in [**Bo11a**], [**Bo11b**].

In the present paper and its continuation [**LieManMar19**] we explore further instances of the relation between cyclotomy and $\mathbb{F}_1$-geometry, including a construction of a λ-ring structure associated to the action of Morse–Smale diffeomorphisms on homology, with roots of unity occurring as eigenvalues. We also extend the Bost–Connes formulation of $\mathbb{F}_1$-geometry through the perspective of motivic measures and of homotopy theoretic spectra. More precisely, we first lift the integral Bost–Connes algebra of [**CCM09**] to the level of a Grothendieck ring of varieties with a good $\hat{\mathbb{Z}}$-action in such a way that the Bost–Connes algebra is obtained from the endomorphisms defined at the level of the Grothendieck ring by taking the equivariant Euler characteristic. Varieties with a good $\hat{\mathbb{Z}}$-action and their associated $\hat{\mathbb{Z}}$-equivariant motives are again a manifestation of $\mathbb{F}_1$-geometry expressed through the presence of cyclotomic structures.

This lift of the Bost–Connes algebra to the $\hat{\mathbb{Z}}$-equivariant Grothendieck ring is then further lifted using the method of assemblers to a homotopy theoretic spectrum. An approach to $\mathbb{F}_1$-geometry via homotopy theoretic spectra and Segal Γ-spaces (which we also use in our construction) was already considered, from a different viewpoint, in [**ToVa09**] and in [**CC16**]. One of our goals in the present investigation is to create a unified framework where cyclotomic points, the Bost–Connes algebra structure, certain motivic zeta functions, and an appropriate construction of spectra, and their various relations to $\mathbb{F}_1$-geometry can be simultaneously realized. In the following part [**LieManMar19**] of this work, we focus on more general motivic measures than the equivariant Euler characteristics we work with here, and we also incorporate the formulation of $\mathbb{F}_1$-structures via torifications introduced in [**LoLo11**] and further developed in [**ManMar16**].

This point of view on lifts of the Bost–Connes formulation of $\mathbb{F}_1$-geometry to Grothendieck rings and spectra is also closely related to the question of categorifications of the Bost–Connes algebra considered in [**MaTa17**]. Categorification is a very broad principle which has been a successfully applied to many different mathematical settings (see [**Mazo12**] for an introductory overview). It is shown in Section 8 of the continuation [**LieManMar19**] of the present paper that the construction based on Grothendieck rings and spectra described here has a parallel version based on Nori motives that maps through a fiber functor, rather than an Euler characteristic, to the motivic categorification of the Bost–Connes algebra constructed in [**MaTa17**]. Our point of view on lifts of the Bost–Connes algebra via motivic measures and zeta functions was further developed in [**LeBr19**].

2. Roots of unity as Weil numbers

2.1. Local zetas and homotopy. In this section we develop the idea sketched in the subsection 0.2 of [**Ma10**]: namely, that a $q = 1$ replacement of the structure $(Fr_q, H^k_{et}(V))$ is a pair $(f_{*k}, H_k(M, \mathbb{Z}))$ where f_{*k} is the action of a Morse–Smale diffeomorphism $f : M \to M$ of a compact manifold M upon its homology $H_k(M, \mathbb{Z})$. In particular, in this model characteristic roots of f_* acting upon (co)homology groups are roots of unity, as might be expected from "Weil numbers in characteristic 1".

We start with basic definitions and references.

2.2. Morse–Smale diffeomorphisms. Below M always means a compact smooth manifold, and $f : M \to M$ its diffeomorphism. A point $x \in M$ is called non–wandering, if for any neighborhood U of x, there is some $n > 0$ such that $U \cap f^n(U) \neq \emptyset$.

Definition 2.1. A function f is called a Morse–Smale diffeomorphism, if

(i) The number of non–wandering points of f is finite.
(ii) f is structurally stable that is, any small variation of f (in the C^r topology) is isotopic to f.

This short definition appears in [**Gr81**]; for a more detailed discussion of geometry, see [**FrSh81**].

2.3. Action of Morse–Smale diffeomorphisms on homology. For any compact manifold M, its homology groups $H_k(M, \mathbb{Z})$ are finitely generated abelian groups. Any diffeomorphism $f : M \to M$ induces automorphisms f_{*k} of these groups. According to [**ShSu75**], if f is Morse–Smale, then each f_{*k} is quasi–unipotent, that is, its eigenvalues are roots of unity.

However, generally this condition is not sufficient. An additional necessary condition is vanishing of a certain obstruction, described in [**FrSh81**] and further studied in [**Gr81**].

Namely, consider the category QI whose objects are pairs (g, H), where H is a (finitely generated) abelian group, and $g : H \to H$ is a quasi–idempotent endomorphism of H (by definition, this means that eigenvalues of g are zero or roots of unity.) Morphisms in QI are self–evident.

The abelian group $K_0(QI)$ admits a morphism onto its torsion subgroup G. Denote by $\varphi : K_0(QI) \to G$ be such a morphism.

Theorem 2.2. [**FrSh81**] *Let $f : M \to M$ be a diffeomorphism such that each $f_{*k} : H_k(M, \mathbb{Z}) \to H_k(M, \mathbb{Z})$ is quasi–unipotent. Let $[f_{*k}]$ be the class in $K_0(QI)$ of the pair $(f_{*k}, H_k(M, \mathbb{Z}))$. Then*

(i) *If f is Morse–Smale, then $\chi(f_*) := \sum_k (-1)^k \varphi([f_{*k}]) \in G$ is zero.*
(ii) *If in addition M is simply connected, and $\dim M > 5$, then vanishing of $\chi(f_*)$ implies that f is isotopic to a Morse–Smale diffeomorphism.*

2.4. Passage from $(f_{*k}, H_k(M,\mathbb{Z}))$ to a scheme over $\mathbb{F}_1$. In this subsection, we sketch a final step from homotopy to $\mathbb{F}_1$–geometry.

One should keep in mind, however, that there might be many divergent paths, starting at this point, because there are several different versions of "geometries over $\mathbb{F}_1$".

We will choose here a version developed in James Borger's paper [**Bo11a**] (and continued in [**Bo11b**]). Roughly speaking, in order to define an affine scheme over $\mathbb{F}_1$, one should give a (commutative) ring with λ–structure and then treat this λ–structure as "descent data" from the base $\operatorname{Spec}\mathbb{Z}$ to the base $\operatorname{Spec}\mathbb{F}_1$.

2.4.1. *From $(f_{*k}, H_k(M,\mathbb{Z}))$ to a λ–ring.* Consider for simplicity only the case (ii) of Theorem 2.2. Then for each k, $H_k(M,\mathbb{Z}))$ is a free $\mathbb{Z}$–module of finite rank, and f_{*k} is its quasi–unipotent automorphism. Fix such a k.

Introduce upon $H_k(M,\mathbb{Z}))$ the structure of $\mathbb{Z}[T,T^{-1}]$–module, where T acts as f_{*k}.

We can consider the minimal subcategory $\mathcal{C}$ of $\mathbb{Z}[T,T^{-1}]$–modules, containing $H_k(M,\mathbb{Z})$ and closed with respect to direct sums, tensor products, and exterior powers, and then produce the Grothendieck λ–ring $K_0(\mathcal{C})$ using exterior powers to define the relevant λ–structure (cf. [**At61**], p. 26, and [**Le81**]).

Definition 2.3. The $\mathbb{F}_1$–scheme, whose Borger's lift to $\operatorname{Spec}\mathbb{Z}$ is $K_0(\mathcal{C})$, is the representative of $(f_{*k}, H_k(M,\mathbb{Z}))$ in $\mathbb{F}_1$–geometry.

Remark 2.4. It seems that another short path from $(f_{*k}, H_k(M,\mathbb{Z}))$ to an $\mathbb{F}_1$–scheme defined differently might lead to one of Le Bruyn's spaces in [**LeBr17**].

2.5. Cases when eigenvalues of conjectural Frobenius maps are not roots of unity. Here we will briefly discuss possible extensions of the picture above, leading to various geometries "below $\operatorname{Spec}\mathbb{Z}$" but generally *not* over $\operatorname{Spec}\mathbb{F}_1$.

The most interesting new virtual zeta–functions of this type were discovered only recently under the generic name "zeta–polynomials", [**Ma16**], [**JMOS16**], [**OnRoSp16**].

In [**Ma16**], it was described how to produce such polynomials from period polynomials of any cusp form over $SGL(2,\mathbb{Z})$ which is an eigenform for all Hecke operators: this passage is a kind of "discrete Mellin transform". It was also proved that zeroes of period polynomials lie on the unit circle of the complex plane, but generally are *not* roots of unity. Both this construction and the results about zeroes were generalised in [**JMOS16**], [**OnRoSp16**] to the case of cusp newforms of even weight for $\Gamma_0(N)$. It turned out that, with appropriate scaling, zeroes of period polynomials lie on the circle $\{z \,|\, |z| = \dfrac{1}{\sqrt{N}}\}$.

Problem 2.5. Make explicit geometries under $\operatorname{Spec}\mathbb{Z}$ in which one can accommodate the respective "motives".

3. The Bost–Connes system and the Grothendieck ring

3.1. $\hat{\mathbb{Z}}$-equivariant Grothendieck ring. We recall the following definitions from [**Lo99**]. Let G be an affine algebraic group acting upon an algebraic variety X. We say that this action is *good*, if each G–orbit is contained in an affine open subset of X.

The Grothendieck ring $K_0^G(\mathcal{V})$ is generated by isomorphism classes $[X]$ of pairs, consisting of varieties with good G–action. Upon these pairs the inclusion–exclusion relations are imposed: $[X] = [Y] + [X \smallsetminus Y]$ where $Y \hookrightarrow X$ and $X \smallsetminus Y \hookrightarrow X$ are G–equivariant embeddings. Multiplication in $K_0^G(\mathcal{V})$ is induced by the diagonal G–action on the product.

In the main special case considered in [**Lo99**],

$$G = \hat{\mathbb{Z}} = \varprojlim_n \mathbb{Z}/n\mathbb{Z}.$$

One defines $K_0^{\hat{\mathbb{Z}}}(\mathcal{V})$ as the Grothendieck ring of varieties with an action of $\hat{\mathbb{Z}}$ that factors through a good action of some finite quotient $\mathbb{Z}/n\mathbb{Z}$.

In this section, we first consider varieties defined over the field $\mathbb{C}$ of complex numbers and classes in the Grothendieck ring correspondingly taken in $K_0(\mathcal{V}_{\mathbb{C}})$ and the equivariant $K_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{C}})$. We then consider the case of varieties defined over $\mathbb{Q}$ with the equivariant Grothendieck ring $K_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{Q}})$. In the first case the target of the equivariant Euler characteristic consists of the abelian part $\mathbb{Z}[\mathbb{Q}/\mathbb{Z}]$ of the integral Bost–Connes algebra, while in the second case it is a subring spanned by the range projectors of the Bost–Connes algebra.

As observed in [**Lo99**], there is an Euler characteristic ring homomorphism

$$\chi^{\hat{\mathbb{Z}}} : K_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{C}}) \to K_0^{\hat{\mathbb{Z}}}(\mathbb{C}),$$

to the Grothendieck ring of finite dimensional representations. Since the character group is $\mathrm{Hom}(\hat{\mathbb{Z}}, \mathbb{G}_m) = \mathbb{Q}/\mathbb{Z}$, the latter is identified with the group ring

$$K_0^{\hat{\mathbb{Z}}}(\mathbb{C}) = \mathbb{Z}[\mathbb{Q}/\mathbb{Z}].$$

3.2. Equivariant Euler characteristics. Since we are considering varieties with a good action of $\hat{\mathbb{Z}}$ that factors through some finite quotient, the Euler characteristic above can be obtained as in [**Ve73**], for an action of a finite group G, by replacing the alternating sum of dimensions of the cohomology groups with a sum in the ring $R(G)$ of representations of G of the classes of cohomology groups, viewed as G-modules.

As observed in [**Gu-Za17**], one can also define an equivariant Euler characteristic, for good actions on varieties of a finite group G, as a ring homomorphism $\chi^G : K_0^G(\mathcal{V}) \to A(G)$, where $A(G)$ is the Burnside ring of G, the Grothendieck ring of the category of finite G-sets. In this case the equivariant Euler characteristic is defined as $\chi^G(X) = \sum_{k\geq 0}[X_k]$ where X has a simplicial decomposition with k–skeleton X_k, such that G acts by simplicial maps which map each k–simplex either identically to itself or to another simplex, so that it makes sense to consider the classes $[X_k]$ in $A(G)$. It is shown in [**Gu-Za17**] that the result is independent of the choice of such a simplicial decomposition. It is also shown that any invariant with values in a commutative ring R, defined on varieties with a good G–action homeomorphic to locally closed unions of cells in finite CW-complexes with G acting by cell maps, which satisfies inclusion-exclusion (on G-invariant decompositions) and multiplicativity on products is necessarily a composition

$$K_0^G(\mathcal{V}) \to A(G) \to R$$

of χ^G with a ring homomorphism $\varphi : A(G) \to R$. In particular, this is the case for the Euler characteristic $K_0^G(\mathcal{V}) \to R(G)$ obtained by composing $\chi^G : K_0^G(\mathcal{V}) \to A(G)$ with the natural ring homomorphism $\varphi : A(G) \to R(G)$ that sends G–sets to their space of functions.

When considering the profinite group $\hat{\mathbb{Z}} = \varprojlim_n \mathbb{Z}/n\mathbb{Z}$, the Burnside rings $A(\mathbb{Z}/n\mathbb{Z})$ form a projective system with limit

$$\hat{A}(\hat{\mathbb{Z}}) = \varprojlim_n A(\mathbb{Z}/n\mathbb{Z}),$$

the completed Burnside ring of $\hat{\mathbb{Z}}$, which is the Grothendieck ring of almost finite $\hat{\mathbb{Z}}$-spaces, namely those $\hat{\mathbb{Z}}$-spaces that are discrete and essentially finite, that is, such that for every open subgroup H the set of points fixed by all elements of H is finite, see Section 2 of [**DrSi88**]. The Burnside ring $A(\hat{\mathbb{Z}})$ of finite $\hat{\mathbb{Z}}$-spaces sits as a dense subring of $\hat{A}(\hat{\mathbb{Z}})$. Moreover, there is an identification of this completed Burnside ring with the Witt ring $\hat{A}(\hat{\mathbb{Z}}) = W(\mathbb{Z})$, see Corollary 1 of [**DrSi88**].

3.3. Lifting the integral Bost–Connes system. Consider now the endomorphisms $\sigma_n : \mathbb{Z}[\mathbb{Q}/\mathbb{Z}] \to \mathbb{Z}[\mathbb{Q}/\mathbb{Z}]$ given by $\sigma_n(e(r)) = e(nr)$ on the standard basis $\{e(r) : r \in \mathbb{Q}/\mathbb{Z}\}$ of $\mathbb{Z}[\mathbb{Q}/\mathbb{Z}]$. The integral Bost–Connes algebra $\mathcal{A}_{\mathbb{Z}}$ introduced in [**CCM09**] is generated by the group ring $\mathbb{Z}[\mathbb{Q}/\mathbb{Z}]$ together with elements $\tilde{\mu}_n$ and μ_n^* satisfying the relations

$$\tilde{\mu}_{nm} = \tilde{\mu}_n \tilde{\mu}_m, \quad \mu_{nm}^* = \mu_n^* \mu_m^*, \quad \mu_n^* \tilde{\mu}_n = n, \quad \tilde{\mu}_n \mu_m^* = \mu_m^* \tilde{\mu}_n, \tag{3.1}$$

where the first two relations hold for arbitrary $n, m \in \mathbb{N}$, the third for arbitrary $n \in \mathbb{N}$ and the fourth for $n, m \in \mathbb{N}$ satisfying $(n, m) = 1$, and the relations

$$x\tilde{\mu}_n = \tilde{\mu}_n \sigma_n(x) \quad \mu_n^* x = \sigma_n(x)\mu_n^*, \quad \tilde{\mu}_n x \mu_n^* = \tilde{\rho}_n(x), \tag{3.2}$$

for any $x \in \mathbb{Z}[\mathbb{Q}/\mathbb{Z}]$, where $\tilde{\rho}_n(e(r)) = \sum_{nr'=r} e(r')$.

The maps $\tilde{\rho}_n$ and the endomorphisms σ_n satisfy the compatibility conditions, for all $x, y \in \mathbb{Z}[\mathbb{Q}/\mathbb{Z}]$ (see Proposition 4.4 of [**CCM09**])

$$\tilde{\rho}_n(\sigma_n(x)y) = x\tilde{\rho}_n(y), \qquad \sigma_n(\tilde{\rho}_m(x)) = (n, m) \cdot \tilde{\rho}_{m'}(\sigma_{n'}(x)), \tag{3.3}$$

where $(n, m) = \gcd\{n, m\}$ and $n' = n/(n, m)$ and $m' = m/(n, m)$.

LEMMA 3.1. *The endomorphisms $\sigma_n : \mathbb{Z}[\mathbb{Q}/\mathbb{Z}] \to \mathbb{Z}[\mathbb{Q}/\mathbb{Z}]$ lift to endomorphisms $\sigma_n : K_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{C}}) \to K_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{C}})$ such that the following diagram commutes*

$$\begin{array}{ccc} K_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{C}}) & \xrightarrow{\chi^{\hat{\mathbb{Z}}}} & \mathbb{Z}[\mathbb{Q}/\mathbb{Z}] \\ \downarrow{\scriptstyle \sigma_n} & & \downarrow{\scriptstyle \sigma_n} \\ K_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{C}}) & \xrightarrow{\chi^{\hat{\mathbb{Z}}}} & \mathbb{Z}[\mathbb{Q}/\mathbb{Z}]. \end{array}$$

These endomorphisms define a semigroup action of the multiplicative semigroup $\mathbb{N}$ on the Grothendieck ring $K_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{C}})$.

PROOF. Let X be a variety with a good $\hat{\mathbb{Z}}$-action, which factors through some finite quotient $\mathbb{Z}/N\mathbb{Z}$. Let $\alpha : \hat{\mathbb{Z}} \times X \to X$ denote the action. The endomorphisms $\sigma_n : \mathbb{Z}[\mathbb{Q}/\mathbb{Z}] \to \mathbb{Z}[\mathbb{Q}/\mathbb{Z}]$ given by $\sigma_n(e(r)) = e(nr)$ have an equivalent description as the action on the group of roots of unity of all orders given by raising to the n–th power $\sigma_n : \zeta \mapsto \zeta^n$. One can then obtain an action $\alpha_n : \hat{\mathbb{Z}} \times X \to X$ given by $\alpha_n = \alpha \circ \sigma_n$. Thus, we assign to a pair (X, α) of a variety with a good $\hat{\mathbb{Z}}$–action the pair (X, α_n) of the same variety with the action α_n. This assignment respects isomorphism classes and is compatible with the relations, hence it determines endomorphisms of $K_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{C}})$, with $\sigma_{nm} = \sigma_n \circ \sigma_m$, namely a semigroup action of the multiplicative semigroup $\mathbb{N}$ on the Grothendieck ring $K_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{C}})$. □

The maps $\tilde{\rho}_n : \mathbb{Z}[\mathbb{Q}/\mathbb{Z}] \to \mathbb{Z}[\mathbb{Q}/\mathbb{Z}]$ of the form

$$\tilde{\rho}_n(e(r)) = \sum_{nr'=r} e(r') \tag{3.4}$$

are not ring homomorphisms but only morphisms of abelian groups. After tensoring with $\mathbb{Q}$, one obtains the group algebra $\mathbb{Q}[\mathbb{Q}/\mathbb{Z}] = \mathbb{Z}[\mathbb{Q}/\mathbb{Z}] \otimes_{\mathbb{Z}} \mathbb{Q}$ on which the $\tilde{\rho}_n$ induce endomorphisms of the form $\rho_n(e(r)) = n^{-1} \sum_{nr'=r} e(r')$ satisfying the relations $\sigma_n \rho_n(x) = x$ and $\rho_n \sigma_n(x) = \pi_n x$, for $x \in \mathbb{Q}[\mathbb{Q}/\mathbb{Z}]$ and the idempotent $\pi_n = n^{-1} \sum_{ns=0} e(s)$. The arithmetic Bost–Connes algebra is the crossed product $\mathbb{Q}[\mathbb{Q}/\mathbb{Z}] \rtimes \mathbb{N}$ generated by $\mathbb{Q}[\mathbb{Q}/\mathbb{Z}]$ and μ_n, μ_n^* with the crossed-product action of $\mathbb{N}$ implemented by $\mu_n x \mu_n^* = \rho_n(x)$, see [**CCM09**].

Once one considers varieties defined over $\mathbb{Q}$, the Grothendieck ring $K_0^{\hat{\mathbb{Z}}}(\mathbb{Q})$ can be characterized as follows.

LEMMA 3.2. *The Grothendieck ring $K_0^{\hat{\mathbb{Z}}}(\mathbb{Q})$ can be identified with the subring of $\mathbb{Z}[\mathbb{Q}/\mathbb{Z}]$ generated by the elements $n\pi_n = \sum_{ns=0} e(s)$.*

PROOF. As noted in [**Lo99**], the element $\sum_{ns=0} e(s)$ in $\mathbb{Z}[\mathbb{Q}/\mathbb{Z}]$ is the image of the irreducible representation of $\mathbb{Z}/n\mathbb{Z}$ given by the cyclotomic field $\mathbb{Q}(\zeta_n)$ seen as a $\mathbb{Q}$-vector space, and these representations give a basis of $K_0^{\hat{\mathbb{Z}}}(\mathbb{Q})$. □

REMARK 3.3. According to the Corollary 4.5 in [**CCM09**], the range of the maps $\tilde{\rho}_n$ in (3.4) is an ideal in $\mathbb{Z}[\mathbb{Q}/\mathbb{Z}]$. This follows from the relation $\tilde{\rho}_n(\sigma_n(x)y) = x\tilde{\rho}_n(y)$ of (3.3). If r' an element of the set $E_n(r) = \{r' \in \mathbb{Q}/\mathbb{Z} : nr' = r\}$, we have $\tilde{\rho}_n(e(r)) = e(r') \sum_{ns=0} e(s)$.

LEMMA 3.4. *There are endomorphisms $\sigma_n : K_0^{\hat{\mathbb{Z}}}(\mathbb{Q}) \to K_0^{\hat{\mathbb{Z}}}(\mathbb{Q})$ induced by the endomorphisms $\sigma_n : \mathbb{Z}[\mathbb{Q}/\mathbb{Z}] \to \mathbb{Z}[\mathbb{Q}/\mathbb{Z}]$. They lift to endomorphisms $\sigma_n : K_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{Q}}) \to K_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{Q}})$ given by $\sigma_n(X, \alpha) = (X, \alpha \circ \sigma_n)$ as in Lemma* 3.1.

PROOF. By identifying $K_0^{\hat{\mathbb{Z}}}(\mathbb{Q})$ with a subring of $\mathbb{Z}[\mathbb{Q}/\mathbb{Z}]$ as in Lemma 3.2, we see that the endomorphisms $\sigma_n : \mathbb{Z}[\mathbb{Q}/\mathbb{Z}] \to \mathbb{Z}[\mathbb{Q}/\mathbb{Z}]$ induce endomorphisms of $K_0^{\hat{\mathbb{Z}}}(\mathbb{Q})$ by the relations $\sigma_n(\tilde{\rho}_m(x)) = (n, m) \cdot \tilde{\rho}_{m'}(\sigma_{n'}(x))$ as in (3.3). Since we have $n\pi_n = \tilde{\rho}_n(1)$, we see that the endomorphisms σ_n map the subring $K_0^{\hat{\mathbb{Z}}}(\mathbb{Q})$ to itself. The lift to $\sigma_n : K_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{Q}}) \to K_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{Q}})$ is obtained by by the same argument as in Lemma 3.1. Namely, the map $\sigma_n(X, \alpha) = (X, \alpha \circ \sigma_n)$ defines an endomorphism of $K_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{Q}})$ which satisfies $\sigma_n \circ \chi^{\hat{\mathbb{Z}}} = \chi^{\hat{\mathbb{Z}}} \circ \sigma_n$, with $\chi^{\hat{\mathbb{Z}}} : K_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{Q}}) \to K_0^{\hat{\mathbb{Z}}}(\mathbb{Q})$ the Euler characteristic. □

In order to lift the maps $\tilde{\rho}_n$ to the level of the Grothendieck ring $K_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{Q}})$, we put

$$\mathbb{V}_n := [Z_n, \gamma_n] \ \in K_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{Q}}),$$

where Z_n is a zero–dimensional variety over $\mathbb{Q}$ with $\#Z_n(\bar{\mathbb{Q}}) = n$. Any such variety with a smooth model over $\operatorname{Spec}\mathbb{Z}$ can be identified with $Z_n = \operatorname{Spec}(\mathbb{Q}^n)$. It is endowed with the natural action $\gamma_n : \hat{\mathbb{Z}} \times Z_n \to Z_n$ that factors through $\mathbb{Z}/n\mathbb{Z}$.

Given a variety X with a good $\hat{\mathbb{Z}}$-action $\alpha : \hat{\mathbb{Z}} \times X \to X$, let

$$\Phi_n(\alpha) : \hat{\mathbb{Z}} \times X \times Z_n \to X \times Z_n$$

be given by

$$\Phi_n(\alpha)(\zeta, x, a_i) = \begin{array}{ll} (x, \gamma_n(\zeta, a_i)) & \text{for} \quad i = 1, \dots, \\ \text{and} \quad (\alpha(\zeta, x), \gamma_n(\zeta, a_n)) & \text{for} \quad i = n. \end{array} \tag{3.5}$$

Notice that (3.5) is just a form of the *Verschiebung* map: for ζ the generator of $\mathbb{Z}/n\mathbb{Z}$ we have

$$\Phi_n(\alpha)(\zeta, x, a_i) = (x, a_{i+1}) \quad \text{for} \quad i = 1, \dots, n-1,$$
$$\text{and} \quad (\alpha(\zeta, x), a_1) \quad \text{for} \quad i = n.$$

Proposition 3.5. *The maps*

$$\tilde{\rho}_n[X, \alpha] := [X \times Z_n, \Phi_n(\alpha)] \tag{3.6}$$

define a homomorphism of the Grothendieck group $K_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{Q}})$ *that satisfies*

$$\sigma_n \circ \tilde{\rho}_n[X, \alpha] = [X, \alpha]^{\oplus n} \tag{3.7}$$

and

$$\tilde{\rho}_n \circ \sigma_n[X, \alpha] = \tilde{\rho}_n[X, \alpha \circ \sigma_n] = [X, \alpha] \cdot [Z_n, \gamma_n]. \tag{3.8}$$

Proof. Given a variety X with a good $\hat{\mathbb{Z}}$-action $\alpha : \hat{\mathbb{Z}} \times X \to X$, consider the product $[X, \alpha] \cdot \mathbb{V}_n$ in $K_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{Q}})$. This class has a representative $[X \times Z_n, (\alpha \times \gamma_n) \circ \Delta]$, where $\Delta : \hat{\mathbb{Z}} \to \hat{\mathbb{Z}} \times \hat{\mathbb{Z}}$ is the diagonal. We have

$$\sigma_n \circ \tilde{\rho}_n[X, \alpha] = \sigma_n[X \times Z_n, \Phi_n(\alpha)] = [X \times Z_n, \Phi_n(\alpha) \circ \sigma_n].$$

We also have

$$\Phi_n(\alpha) \circ \sigma_n = (\alpha \times 1) \circ \Delta,$$

since

$$\Phi_n(\alpha) \circ \sigma_n(\zeta, x, a_i) = \Phi_n(\alpha)(\zeta^n, x, a_i) = (\Phi_n(\alpha)(\zeta))^n(x, a_i),$$

where we write $\Phi_n(\alpha)(\zeta) : X \times Z_n \to X \times Z_n$ for the action of $\zeta \in \hat{\mathbb{Z}}$, with $\Phi_n(\alpha)(\zeta_1 \cdots \zeta_n) = \Phi_n(\alpha)(\zeta_1) \circ \cdots \circ \Phi_n(\alpha)(\zeta_n)$, and the n-fold composition gives

$$\Phi_n(\alpha)(\zeta) \circ \cdots \circ \Phi_n(\alpha)(\zeta)(x, a_i) = (\alpha(\zeta, x), a_i).$$

This shows (3.7).

The second relation is obtained similarly. We have

$$\tilde{\rho}_n \circ \sigma_n[X, \alpha] = \tilde{\rho}_n[X, \alpha \circ \sigma_n] = [X \times Z_n, \Phi_n(\alpha \circ \sigma_n)],$$

where

$$\Phi_n(\alpha)(\zeta, x, a_i) = (x, a_{i+1}) \quad \text{for} \quad i = 1, \ldots, n-1,$$
$$\text{and} \quad (\alpha(\zeta^n, x), a_1) \quad \text{for} \quad i = n.$$

Now $\alpha(\zeta^n, x) = \alpha(\zeta)^n(x)$, hence we have $\Phi_n(\alpha \circ \sigma_n)(\zeta)(x, a_i) = \Phi_n(\alpha(\zeta)^n)(x, a_i)$. The usual relations $V_n(F_n(a)b) = aV_n(b)$ between Frobenius F_n and Verschebung V_n (see Proposition 2.2 of [**CC14**]) holds in this case as well in the form $\Phi_n(\alpha(\zeta)^n) = \alpha(\zeta)\Phi_n(1)$ where $\Phi_n(1)(x, a_i) = (x, a_{i+1})$ is the cyclic permutation action of $\mathbb{Z}/n\mathbb{Z}$ on Z_n. Thus, we obtain $\Phi_n(\alpha \circ \sigma_n) = (\alpha \times \gamma_n) \circ \Delta$. This gives (3.8). □

The relation (3.7) corresponds to $\sigma_n \circ \tilde{\rho}_n(x) = nx$, and (3.8) to $\tilde{\rho}_n \circ \sigma_n(x) = n\pi_n x$, for $x \in \mathbb{Z}[\mathbb{Q}/\mathbb{Z}]$. They are geometric manifestations of the same relation between the maps σ_n and $\tilde{\rho}_n$ of the integral Bost–Connes system and the Frobenius and Verschiebung described in [**CC14**]. For other occurrences of the same relation see also [**MaRe17**], [**MaTa17**].

3.4. A non–commutative extension of the Grothendieck ring. Let $\mathbf{K}_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{Q}})$ be the non-commutative ring generated by $K_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{Q}})$ and elements $\tilde{\mu}_n$, μ_n^* for $n \in \mathbb{N}$ satisfying the relations (3.1) for all $n, m \in \mathbb{N}$, and (3.2) for all $x = [X, \alpha] \in K_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{Q}})$ and all $n \in \mathbb{N}$.

Lemma 3.6. *The Euler characteristic $\chi^{\hat{\mathbb{Z}}} : K_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{Q}}) \to K_0^{\hat{\mathbb{Z}}}(\mathbb{Q}) \hookrightarrow \mathbb{Z}[\mathbb{Q}/\mathbb{Z}]$ extends to a ring homomorphism $\chi : \mathbf{K}_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{Q}}) \to \mathcal{A}_{\mathbb{Z}}$ to the integral Bost–Connes algebra. After tensoring with $\mathbb{Q}$, we obtain a homomorphism of semigroup crossed product rings*

$$\chi : K_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{Q}})_{\mathbb{Q}} \rtimes \mathbb{N} \to \mathbb{Q}[\mathbb{Q}/\mathbb{Z}] \rtimes \mathbb{N},$$

where $\mathbf{K}_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{Q}}) \otimes_{\mathbb{Z}} \mathbb{Q} = K_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{Q}})_{\mathbb{Q}} \rtimes \mathbb{N}$ with $K_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{Q}})_{\mathbb{Q}} = K_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{Q}}) \otimes_{\mathbb{Z}} \mathbb{Q}$, and $\mathcal{A}_{\mathbb{Z}} \otimes_{\mathbb{Z}} \mathbb{Q} = \mathcal{A}_{\mathbb{Q}} = \mathbb{Q}[\mathbb{Q}/\mathbb{Z}] \rtimes \mathbb{N}$ is the rational Bost–Connes algebra.

Proof. We define the map χ as $\chi^{\hat{\mathbb{Z}}}$ on elements of $K_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{Q}})$ and the identity on the extra generators $\chi(\tilde{\mu}_n) = \tilde{\mu}_n$ and $\chi(\mu_n^*) = \mu_n^*$. By Lemma 3.4 and Proposition 3.5, this map is compatible with the relations in $\mathbf{K}_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{Q}})$ and in $\mathcal{A}_{\mathbb{Z}}$. After tensoring with $\mathbb{Q}$, the algebra $\mathbf{K}_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{Q}}) \otimes_{\mathbb{Z}} \mathbb{Q}$ can be identified with a semigroup crossed product by taking as generators the elements of $K_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{Q}})$ and $\mu_n = n^{-1}\tilde{\mu}_n$ and μ_n^*, which satisfy the relations

$$\mu_n^*\mu_n = 1, \quad \mu_{nm} = \mu_n\mu_m, \quad \mu_{nm}^* = \mu_n^*\mu_m^*, \quad \forall n, m \in \mathbb{N},$$
$$\mu_n\mu_m^* = \mu_m^*\mu_n \text{ if } (n,m) = 1$$
$$\mu_n x \mu_n^* = \rho_n(x) \quad \text{with} \quad \rho_n(x) = \frac{1}{n}\tilde{\rho}_n(x),$$

with $\sigma_n\rho_n(x) = x$, for all $x = [X, \alpha] \in K_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{Q}})$. The semigroup action in the crossed product $K_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{Q}})_{\mathbb{Q}} \rtimes \mathbb{N}$ is given by $x \mapsto \rho_n(x) = \mu_n x \mu_n^*$. The target algebra is the rational Bost–Connes algebra $\mathcal{A}_{\mathbb{Z}} \otimes_{\mathbb{Z}} \mathbb{Q} = \mathbb{Q}[\mathbb{Q}/\mathbb{Z}] \rtimes \mathbb{N}$. Again the map given by $\chi^{\hat{\mathbb{Z}}}$ on elements of $K_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{Q}})$ and by $\chi(\mu_n) = \mu_n$ and $\chi(\mu_n^*) = \mu_n^*$ determines a homomorphism of crossed-product algebras. □

4. From Rings to Spectra

4.1. Spectra. We give a brief review of spectra, with the purpose of recalling a construction of Segal [**Se74**] that constructs spectra from Gamma spaces. We then review the notion of assembler categories [**Za17a**] and how they can be used to construct a Gamma space and an associated spectrum whose π_0 realizes certain abstract scissor-congruence relations.

The construction of spectra from Gamma spaces was first developed in the context of the Bousfield–Friedlander spectra, see Definition 2.1 of [**BousFr78**].

In this setting, one considers the simplicial category Δ, which has an object $[n]$ for each $n \in \mathbb{N}$ given by the finite totally ordered set $[n] = \{0 < 1 < \ldots < n-1\}$, with morphisms the face and degeneracy maps δ_i^n and σ_i^n satisfying the simplicial relations.

A simplicial object is a functor $S : \Delta^{op} \to \mathcal{C}$ from the simplicial category of a given category $\mathcal{C}$. It is determined by a sequence of objects $X(n)$ of $\mathcal{C}$ with morphisms corresponding to faces and degeneracies. We denote by $\Delta(\mathcal{C})$ the resulting category of simplicial objects in $\mathcal{C}$. In particular, a simplicial set is a simplicial object in the category of sets and we will use the notation $\underline{\Delta} = \Delta(\text{Sets})$ for the category of simplicial sets.

Similarly, a bisimplicial object is a functor $BS : \Delta^{op} \times \Delta^{op} \to \mathcal{C}$, or equivalently a simplicial object in the category of simplicial objects $\Delta(\mathcal{C})$. The diagonal of a bisimplicial object BS is the simplicial object obtained by precomposition of BS with the diagonal functor $\Delta^{op} \to \Delta^{op} \times \Delta^{op}$.

The category $\mathcal{S}$ of Bousfield–Friedlander spectra has objects X given by sequences of simplicial sets $X = \{X_n\}_{n \geq 0}$ endowed with structure maps $\varphi_n^X : S^1 \wedge X_n \to X_{n+1}$ for all $n \geq 0$, and morphisms given by maps $f_n : X_n \to Y_n$ with commutative diagrams

$$\begin{array}{ccc} S^1 \wedge X_n & \xrightarrow{\varphi_n^X} & X_{n+1} \\ \big\downarrow{\scriptstyle 1_{S^1} \wedge f_n} & & \big\downarrow{\scriptstyle f_{n+1}} \\ S^1 \wedge Y_n & \xrightarrow{\varphi_n^Y} & Y_{n+1}. \end{array}$$

The sphere spectrum $\mathbb{S}$ has $\mathbb{S}_n = S^1 \wedge \cdots \wedge S^1$, the n–fold smash product, and φ_n the identity map.

Let $\gamma_n^X : X_n \to \Omega X_{n+1}$ be the maps induced by the adjoints of the structure maps. An Ω-spectrum is a spectrum where the maps γ_n^X are weak equivalences for all n.

The homotopy groups $\pi_k(X)$ of spectra are given by

$$\pi_k(X) = \varinjlim_n \pi_{k+n}(X_n)$$

over the maps $\pi_{k+n}(X_n) \to \pi_{k+n}(\Omega X_{n+1}) \simeq \pi_{n+k+1}(X_{n+1})$, induced the γ_n^X. A spectrum is n–connected if $\pi_k(X) = 0$ for all $k \leq n$ and connective if it is -1-connected. A spectrum X is cofibrant if all the structure maps $\phi_n^X : S^1 \wedge X_n \to X_{n+1}$ are cofibrations.

However, a problem with the Bousfield–Friedlander spectra is that they do not have a homotopically good smash product. The construction of categories of spectra with smash products were developed in the '90s, especially the S-modules model of [**EKMM97**] and the symmetric spectra model of [**HSS00**]. In a more modern approach, it is therefore preferable to work with symmetric spectra for the Segal construction. Indeed, the Gamma-spaces, SW-categories and Waldhausen categories that occur in relation to the spectra underlying the Grothendieck ring of varieties and its variants naturally give rise to symmetric spectra.

A symmetric spectrum consists of a sequence of pointed spaces (pointed simplicial sets) $X = \{X_n\}_{n\geq 0}$ together with a left action of the symmetric group S_n on X_n for all $n \geq 0$ and structure maps given by based maps $\varphi_n^X : S^1 \wedge X_n \to X_{n+1}$ for all $n \geq 0$, with the condition that, for all $n, m \geq 0$ the composition $\varphi^X_{n+m-1} \circ \cdots \circ \varphi^X_n$

$$\varphi^X_{n+m-1} \circ \cdots \circ \varphi^X_n : S^m \wedge X_n \to S^{m-1} \wedge X_{n+1} \to \cdots S^1 \wedge X_{n+m-1} \to X_{n+m}$$

is $S_n \times S_m$-equivariant. A morphism of symmetric spectra is a collection of S_n-equivariant based maps $f_n : X_n \to Y_n$ such that $f_{n+1} \circ \varphi^X_n = \varphi^Y_n \circ (f_n \wedge Id_{S^1})$, for all $n \geq 0$. A symmetric spectrum is ring spectrum if it is also endowed with $S_n \times S_m$-equivariant multiplication maps

$$M_{n,m} : X_n \wedge X_m \to X_{n+m}$$

and unit maps $\iota_0 : S^0 \to X_0$ and $\iota_1 : S^1 \to X_1$ satisfying the associativity commutative squares

$$\begin{array}{ccc} X_n \wedge X_m \wedge X_r & \xrightarrow{Id \wedge M_{m,r}} & X_n \wedge X_{m+r} \\ {\scriptstyle M_{n,m}\wedge Id}\downarrow & & \downarrow{\scriptstyle M_{n,m+r}} \\ X_{n+m} \wedge X_r & \xrightarrow{M_{n+m,r}} & X_{n+m+r} \end{array}$$

the unit relations

$$M_{n,0} \circ (Id \wedge \iota_0) = Id : X_n \simeq X_n \wedge S^0 \to X_n \wedge X_0 \to X_n$$

and similarly $M_{0,n} \circ (\iota_0 \wedge Id) = Id$ for all $n \geq 0$, as well as $\chi_{n,1} \circ (M_{n,1} \circ (Id \wedge \iota_1)) = (M_{1,n} \circ (\iota_1 \wedge Id)) \circ \tau$ with $\tau : X_n \wedge S^1 \to S^1 \wedge X_n$ and $\chi_{n,m} \in S_{n+m}$ the shuffle permutation moving the first n elements past the last m. Commutativity of a symmetric ring spectrum is expressed by the commutativity of the diagrams

$$\begin{array}{ccc} X_n \wedge X_m & \longrightarrow & X_m \wedge X_n \\ {\scriptstyle M_{n,m}}\downarrow & & \downarrow{\scriptstyle M_{m,n}} \\ X_{n+m} & \xrightarrow[\chi_{n,m}]{} & X_{m+n} \end{array}$$

with the twist as the first map. For a detailed introduction to symmetric spectra we refer the reader to [**Schw12**].

4.2. Γ–spaces. We recall the setting of Γ–spaces used in the Segal's construction of spectra from categorical data. The notion of Γ-spaces and its relation to connective spectra formalizes the intuition that spectra are a natural homotopy-theoretic generalization of abelian groups.

Let Γ^0 denote the category of finite pointed sets, with objects

$$\underline{n} = \{0, 1, 2, \ldots, n\}$$

and morphisms $f \in \Gamma^0(\underline{n}, \underline{m})$ given by functions

$$f : \{0,1,2,\dots,n\} \to \{0,1,2,\dots,m\}, \quad \text{with} \;\; f(0) = 0.$$

Let Γ denote the opposite category.

A pointed category $\mathcal{C}$ is a category with a chosen object that is both initial and final. A pointed functor $F : \Gamma^0 \to \mathcal{C}$ is called a Γ-object in $\mathcal{C}$.

Given a pointed category $\mathcal{C}$, the category $\Gamma\mathcal{C}$ has objects the pointed functors $F : \Gamma^0 \to \mathcal{C}$ and morphisms the natural transformations between these functors.

Γ-spaces are objects of the category $\Gamma\mathcal{C}$, in the case where $\mathcal{C} = \underline{\Delta}_*$ is the category of pointed simplicial sets.

Given a Γ-space $F : \Gamma^0 \to \underline{\Delta}_*$, the morphisms $f_j : \underline{n} \to \underline{1}$ that map the j-th element to 1 and the rest to 0 induce, for each $n \geq 1$, a morphism

$$F(\underline{n}) \to \prod_{j=1}^{n} F(\underline{1}). \tag{4.1}$$

The *special* Γ–spaces (or Segal Γ–spaces) are Γ–spaces F as above, where all the maps (4.1) are weak equivalences. For special Γ–spaces the weak equivalence $F(\underline{2}) \simeq F(\underline{1}) \times F(\underline{1})$ induces a monoid

$$\pi_0(F(\underline{1})) \times \pi_0(F(\underline{1})) \to \pi_0(F(\underline{2})) \to \pi_0(F(\underline{1})).$$

Such a Γ–space is called *very special* when this monoid is an abelian group.

The Γ–space $\mathbb{S} : \Gamma^0 \to \underline{\Delta}_*$ is given by the inclusion of the category Γ^0 into $\underline{\Delta}_*$ mapping a finite pointed set to the corresponding discrete pointed simplicial set. As shown in [**Se74**] (Barratt–Priddy–Quillen theorem), the associated spectrum is the sphere spectrum, which we also denote by $\mathbb{S}$.

The category Γ^0 of finite pointed sets has a smash product functor $\wedge : \Gamma^0 \times \Gamma^0 \to \Gamma^0$, with $(\underline{n}, \underline{m}) \mapsto \underline{n} \wedge \underline{m}$, which extends to a smash product of arbitrary pointed (simplicial) sets.

The smash product of Γ-spaces constructed in [**Ly99**] is obtained by first associating to a pair $F, F' : \Gamma^0 \to \underline{\Delta}_*$ of Γ–spaces a bi–Γ-space $F \tilde{\wedge} F' : \Gamma^0 \times \Gamma^0 \to \underline{\Delta}_*$

$$(F \tilde{\wedge} F')(\underline{n}, \underline{m}) = F(\underline{n}) \wedge F'(\underline{m})$$

and then defining

$$(F \wedge F')(\underline{n}) = \mathrm{colim}_{\underline{k} \wedge \underline{\ell} \to \underline{n}} (F \tilde{\wedge} F')(\underline{k}, \underline{\ell}),$$

where $\underline{k} \wedge \underline{\ell}$ is the smash product $\wedge : \Gamma^0 \times \Gamma^0 \to \Gamma^0$. It is shown in [**Ly99**] that, up to natural isomorphism, this smash product is associative and commutative and with unit given by the Γ-space $\mathbb{S}$, and that the category of Γ-spaces is symmetric monoidal with respect to this product.

Another use of Γ-spaces in the context of $\mathbb{F}_1$-geometry can be found in the recent paper [**CC16**].

4.3. From Γ–spaces to connective spectra. The construction of [**Se74**], and more generally [**BousFr78**], assigns a connective spectrum to a Γ-space in such a way as to obtain an equivalence between the homotopy category of Γ-spaces and the homotopy category of connective spectra. The construction of spectra from Γ-spaces can be performed in the modern setting of symmetric spectra, rather than in the original Bousfield–Friedlander formulation of [**BousFr78**], see Chapter I, Section 7.4 of [**Schw12**].

If X is a simplicial set, one denotes by X_* the pointed simplicial set obtained by adding a disjoint base point. Given a Γ-space $F : \Gamma^0 \to \underline{\Delta}_*$ and a pointed simplicial set X, one obtains a new Γ-space $X \wedge F$, which maps $\underline{n} \in \Gamma^0$ to $X \wedge F(\underline{n})$ in $\underline{\Delta}_*$.

Recall that, given a functor $F : \mathcal{C}^{op} \times \mathcal{C} \to \mathcal{D}$, the coend $\int^{C \in \mathcal{C}} F(C, C)$ is the initial cowedge, where a cowedge to an object X in $\mathcal{C}$ is a family of morphisms $h_A : A \to X$, for each $A \in \mathcal{C}$, such that, for all morphisms $f : A \to B$ in $\mathcal{C}$ the following diagrams commute

$$\begin{array}{ccc} F(B,A) & \xrightarrow{F(f,A)} & F(A,A) \\ \downarrow{\scriptstyle F(B,f)} & & \downarrow{\scriptstyle h_A} \\ F(B,B) & \xrightarrow{h_B} & X. \end{array}$$

The key step in the construction of a connective spectrum associated to a Γ–space consists of extending a Γ–space F to an endofunctor of the category of pointed simplicial sets. This endofunctor is defined in [**BousFr78**] (see also [**Schw99**]) as the functor (still denoted by F) that maps a pointed simpliciat set $K \in \underline{\Delta}_*$ to the coend

$$F : K \mapsto \int^{\underline{n} \in \Gamma^{op}} K^n \wedge F(\underline{n}),$$

with natural assembly maps $K \wedge F(K') \to F(K \wedge K')$.

The spectrum associated to F, which we denote by $F(\mathbb{S})$, is then given by the sequence of pointed simplicial sets $F(\mathbb{S})_n = F(S^n = S^1 \wedge \cdots \wedge S^1)$, with the maps $S^1 \wedge F(S^n) \to F(S^{n+1})$.

The smash product of Γ-spaces is compatible with the smash product of spectra: as shown in [**Ly99**], if $F, F' : \Gamma^0 \to \underline{\Delta}_*$ are Γ-spaces with $F(\mathbb{S})$ and $F'(\mathbb{S})$ the corresponding spectra, then there is a map of spectra,

$$F(\mathbb{S}) \wedge F'(\mathbb{S}) \to (F \wedge F')(\mathbb{S})$$

which is natural in (F, F'), and a weak equivalence if one of the factors is cofibrant.

This gives rise to a notion of ring spectra (see [**Schw99**]) defined as the monoids in the symmetric monoidal category of Γ–spaces with the smash product of [**Ly99**] recalled above. One refers to these as Γ–rings. Namely, a Γ–ring is a Γ–space F endowed with unit and multiplication maps $\mathbb{S} \to F$ and $F \wedge F \to F$ with associativity and unit properties (Sec VII.3 of [**MacL71**]). The associated connective spectrum of a commutative Γ-ring is a commutative symmetric ring spectrum. However, not all connective commutative symmetric ring spectra come from a commutative Γ–ring, see [**La09**]. For a comparative view of the settings of Γ-rings and symmetric ring spectra, see the discussion in Section 2 of [**Schw99**].

If G is an abelian group, there is an associated Γ-space HG given on objects by

$$HG(\underline{n}) = G \otimes \mathbb{Z}[\underline{n}] \simeq G^n,$$

where $\mathbb{Z}[\underline{n}]$ is the free abelian group on the finite set $\underline{n}$. If $f : \underline{n} \to \underline{m}$ is a morphism in Γ^0, then the associated morphism $H(f) : HG(\underline{n}) \to HG(\underline{m})$ maps an n-tuple $(g_1, \dots, g_n)$ (with $g_0 = 0$) in G^n to the m–tuple $(\sum_{j \in f^{-1}(1)} g_j, \dots, \sum_{j \in f^{-1}(n)} g_n)$. This Eilenberg–MacLane Γ–space HG maps to the Eilenberg–MacLane spectrum of G, which we still denote by HG. If R is a simplicial ring, then HR is an $\mathbb{S}$–algebra with multiplication $HR \wedge HR \to H(R \otimes R) \to HR$ and unit $\mathbb{S} \to H\mathbb{Z} \to HR$.

4.4. Assemblers, spectra, and the Grothendieck ring. We start with a brief survey of the construction of a spectrum associated to the Grothendieck ring of varieties developed in [**Za17a**] and [**Za17c**].

Inna Zakharevich developed in [**Za17a**] and [**Za17b**] a very general formalism for scissor–congruence relations. The abstract form of scissor–congruence consists of categorical data called *assemblers*, which in turn determine a homotopy–theoretic *spectrum*, whose homotopy groups embody scissor–congruence relations. This formalism is applied in [**Za17c**] in the framework producing an assembler and a spectrum whose π_0 recovers the Grothendieck ring of varieties. This is used to obtain a characterisation of the kernel of multiplication by the Lefschetz motive, which provides a general explanation for the observations of [**Bor14**], [**Mart16**] on the fact that the Lefschetz motive is a zero divisor in the Grothendieck ring of varieties.

A *sieve* in a category $\mathcal{C}$ is a full subcategory $\mathcal{C}'$ that is closed under precomposition by morphisms in $\mathcal{C}$. A *Grothendieck topology* on a category $\mathcal{C}$ consists of the assignment of a collection $\mathcal{J}(X)$ of sieves in the over category $\mathcal{C}/X$, for each object X in $\mathcal{C}$, with the following properties:

(i) the over category $\mathcal{C}/X$ is in the collection $\mathcal{J}(X)$;

(ii) the pullback of a sieve in $\mathcal{J}(X)$ under a morphism $f : Y \to X$ is a sieve in $\mathcal{J}(Y)$;

(iii) given $\mathcal{C}' \in \mathcal{J}(X)$ and a sieve $\mathcal{T}$ in $\mathcal{C}/X$, if for every $f : Y \to X$ in $\mathcal{C}'$ the pullback $f^*\mathcal{T}$ is in $\mathcal{J}(Y)$ then $\mathcal{T}$ is in $\mathcal{J}(X)$.

Let $\mathcal{C}$ be a category with a Grothendieck topology. A collection of morphisms $\{f_i : X_i \to X\}_{i \in I}$ in $\mathcal{C}$ is a *covering family* if the full subcategory of $\mathcal{C}/X$ that contains all the morphisms of $\mathcal{C}$ that factor through the f_i,

$$\{g : Y \to X \,|\, \exists i \in I \; h : Y \to X_i \text{ such that } f_i \circ h = g\},$$

is in the sieve collection $\mathcal{J}(X)$.

In a category $\mathcal{C}$ with an initial object $\emptyset$ two morphisms $f : Y \to X$ and $g : W \to X$ are called *disjoint* if the pullback $Y \times_X W$ exists and is equal to $\emptyset$. A collection $\{f_i : X_i \to X\}_{i \in I}$ in $\mathcal{C}$ is disjoint if f_i and f_j are disjoint for all $i \neq j \in I$.

An *assembler category* $\mathcal{C}$ is a small category endowed with a Grothendieck topology, which has an initial object $\emptyset$ (with the empty family as covering family), and where all morphisms are monomorphisms, with the property that any two finite disjoint covering families of X in $\mathcal{C}$ have a common refinement that is also a finite disjoint covering family.

A morphism of assemblers is a functor $F : \mathcal{C} \to \mathcal{C}'$ that is continuous for the Grothendieck topologies and preserves the initial object and the disjointness property, that is, if two morphisms are disjoint in $\mathcal{C}$ their images are disjoint in $\mathcal{C}'$.

For X a finite set, the coproduct of assemblers $\bigvee_{x\in X} \mathcal{C}_x$ is a category whose objects are the initial object $\emptyset$ and all the non–initial objects of the assemblers $\mathcal{C}_x$. Morphisms of non–initial objects are induced by those of $\mathcal{C}_x$.

The abstract scissor congruences consist of pairs of an assembler $\mathcal{C}$ and a sieve $\mathcal{D}$ in $\mathcal{C}$. Given such a pair, one has an associated assembler $\mathcal{C} \smallsetminus \mathcal{D}$ defined as the full subcategory of $\mathcal{C}$ that contains all the objects that are not non–initial objects of $\mathcal{D}$. The assembler structure on $\mathcal{C} \smallsetminus \mathcal{D}$ is determined by taking as covering families in $\mathcal{C} \smallsetminus \mathcal{D}$ those collections $\{f_i : X_i \to X\}_{i\in I}$ with X_i, X objects in $\mathcal{C} \smallsetminus \mathcal{D}$ that can be completed to a covering family in $\mathcal{C}$, namely such that there exists $\{f_j : X_j \to X\}_{j\in J}$ with X_j in $\mathcal{D}$ such that $\{f_i : X_i \to X\}_{i\in I} \cup \{f_j : X_j \to X\}_{j\in J}$ is a covering family in $\mathcal{C}$. There is a morphism of assemblers $\mathcal{C} \to \mathcal{C} \smallsetminus \mathcal{D}$ that maps objects of $\mathcal{D}$ to $\emptyset$ and objects of $\mathcal{C} \smallsetminus \mathcal{D}$ to themselves and morphisms with source in $\mathcal{C} \smallsetminus \mathcal{D}$ to themselves and morphisms with source in $\mathcal{D}$ to the unique morphism to the same target with source $\emptyset$. The data $\mathcal{C}, \mathcal{D}, \mathcal{C} \smallsetminus \mathcal{D}$ are an *abstract scissor congruence*, [**Za17a**], [**Za17b**].

The construction of spectra from assembler categories uses the general construction of spectra from categorical data is provided by the Segal construction [**Se74**] of spectra from Γ–spaces, that we recalled in Section 4.2 above.

The main construction of [**Za17a**] associates to an assembler $\mathcal{C}$ a homotopy–theoretic spectrum, whose homotopy groups provide a family of associated topological invariants satisfying versions of scissor congruence relations. The main steps of the construction can be summarized as follows (see [**Za17a**]):

(1) One associates to an assembler $\mathcal{C}$ a category $\mathcal{W}(\mathcal{C})$ with objects $\{A_i\}_{i\in I}$, collections of non–initial objects A_i of $\mathcal{C}$ indexed by a finite set I, and morphisms $f : \{A_i\}_{i\in I} \to \{B_j\}_{j\in J}$ given by a map of finite sets $f : I \to J$ and morphisms $f_i : A_i \to B_{f(i)}$ such that $\{f_i : A_i \to B_j \,:\, i \in f^{-1}(j)\}$ is a disjoint covering family for all $j \in J$.

(2) For a finite pointed set (X, x_0) and an assembler $\mathcal{C}$, one considers the assembler $X \wedge \mathcal{C} := \bigvee_{x\in X\smallsetminus\{x_0\}} \mathcal{C}$. The assignment $X \mapsto \mathcal{N}\mathcal{W}(X \wedge \mathcal{C})$, where $\mathcal{N}$ is the nerve, is a Γ–space in the sense of [**Se74**] recalled above, hence it defines a spectrum $K(\mathcal{C})$ by

$$X_n = \mathcal{N}\mathcal{W}(S^n \wedge \mathcal{C})$$

with structure maps $S^1 \wedge X_n \to X_{n+1}$ determined by the maps

$$X \wedge \mathcal{N}\mathcal{W}(\mathcal{C}) \to \mathcal{N}\mathcal{W}(X \wedge \mathcal{C}).$$

(3) The group $K_0(\mathcal{C}) := \pi_0 K(\mathcal{C})$ is the free abelian group generated by objects of $\mathcal{C}$ modulo the scissor–congruence relations $[A] = \sum_{i\in I}[A_i]$ for each finite disjoint covering family $\{A_i \to A\}_{i\in I}$.

(4) Given a morphism $\varphi : \mathcal{C}_1 \to \mathcal{C}_2$ of assemblers, there is an assembler $\mathcal{C}_2/\varphi$ and a morphism $\iota : \mathcal{C}_2 \to \mathcal{C}_2/\varphi$ such that the diagram

$$K(\iota) \circ K(\varphi) : \; K(\mathcal{C}_1) \to K(\mathcal{C}_2) \to K(\mathcal{C}_2/\varphi)$$

is a cofiber sequence.

4.5. Assembler for the equivariant Grothendieck ring. As we have seen in the previous section, the equivariant version of the Grothendieck ring $K_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{K}})$ is generated by isomorphism classes of varieties with a "good" $\hat{\mathbb{Z}}$-action, where as before good means that each orbit is contained in an affine open subvariety of X and that the action factors through some finite level $\mathbb{Z}/N\mathbb{Z}$. The scissor congruence relations in $K_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{K}})$ are of the form $[X] = [Y] + [X \smallsetminus Y]$ where $Y \hookrightarrow X$ and $X \smallsetminus Y \hookrightarrow X$ are $\hat{\mathbb{Z}}$-equivariant embeddings. The product is given by the Cartesian product endowed with the induced diagonal $\hat{\mathbb{Z}}$-action.

LEMMA 4.1. *The category $\mathcal{C}^{\hat{\mathbb{Z}}}$ with objects that are varieties X with a good $\hat{\mathbb{Z}}$-action and morphisms that are equivariant locally closed embeddings, endowed with the Grothendieck topology generated by the covering families $\{Y \hookrightarrow X, X \smallsetminus Y \hookrightarrow X\}$ of $\hat{\mathbb{Z}}$-equivariant embeddings, is an assembler category. The spectrum $K^{\hat{\mathbb{Z}}}(\mathcal{V})$ determined by the assembler $\mathcal{C}^{\hat{\mathbb{Z}}}$ has π_0 given by the equivariant Grothendieck ring $K_0^{\hat{\mathbb{Z}}}(\mathcal{V})$.*

PROOF. The first part of the statement follows as in Example 1 of sec. 1 of [**Za17c**]. The empty set is the initial object. Finite disjoint covering families are $\hat{\mathbb{Z}}$-equivariant maps $f_i : X_i \hookrightarrow X$ where $X_i = Y_i \smallsetminus Y_{i-1}$ for a chain of $\hat{\mathbb{Z}}$-equivariant embeddings $\emptyset = Y_0 \hookrightarrow Y_1 \hookrightarrow \cdots \hookrightarrow Y_n = X$. The property that any two finite disjoint covering families have a common refinement follows since the category has pullbacks, [**Za17a**]. Morphisms are compositions of closed and open $\hat{\mathbb{Z}}$-equivariant embeddings, hence they are all monomorphisms. For the second part, by Theorem 2.3 of [**Za17a**], if $K(\mathcal{C})$ is the spectrum determined by an assembler $\mathcal{C}$, then $\pi_0 K(\mathcal{C})$ is generated, as an abelian group, by the objects of $\mathcal{C}$ with the scissor–congruence relations determined by disjoint covering families. In this case this means that $K(\mathcal{C}^{\hat{\mathbb{Z}}})$ is generated by the pairs (X, α) of a variety X with a good $\hat{\mathbb{Z}}$-action α with relations $[X] = [Y] + [X \smallsetminus Y]$ for the covering families given by $\hat{\mathbb{Z}}$-equivariant embeddings $\{Y \hookrightarrow X, X \smallsetminus Y \hookrightarrow X\}$. The ring structure is coming from the symmetric monoidal structure on the category of assemblers, which induces an E_∞–ring structure on the spectrum $K(\mathcal{C})$. In this case it induces the product on $K_0^{\hat{\mathbb{Z}}}(\mathcal{V})$ given by the Cartesian product of varieties with the diagonal $\hat{\mathbb{Z}}$-action (see also Theorem 1.4 of [**Ca15**]). The ring structure is induced by an E_∞–ring spectrum structure on $K(\mathcal{C})$ which is in turn induced by a symmetric monoidal structure on the category of assembler, cf. [**Za17a**]. □

Following Theorem 4.25 of [**Ca15**], the ring structure on $K_0^{\hat{\mathbb{Z}}}(\mathcal{V})$ can also be seen, as in the case of the ordinary Grothendieck ring $K_0(\mathcal{V})$, as induced on π_0 by an E_∞–ring spectrum structure obtained from the fact that the cartesian product of varieties determines a biexact symmetric monoidal structure on $\mathcal{V}$, seen as an SW–category (a subtractive Waldhausen category).

4.6. Lifting the Bost–Connes algebra to spectra. We will now show how to lift the maps σ_n and $\tilde{\rho}_n$ of the Bost–Connes system from the level of the Grothendieck ring $K_0^{\hat{\mathbb{Z}}}(\mathcal{V})$ to the level of the spectrum $K^{\hat{\mathbb{Z}}}(\mathcal{V})$.

PROPOSITION 4.2. *The maps $\sigma_n(X,\alpha) = (X, \alpha \circ \sigma_n)$ and $\tilde{\rho}_n(X,\alpha) = (X \times Z_n, \Phi_n(\alpha))$ on $K_0^{\hat{\mathbb{Z}}}(\mathcal{V})$ determine endofunctors of the assembler category $\mathcal{C}^{\hat{\mathbb{Z}}}$. The endofunctors σ_n are compatible with the monoidal structure induced by the Cartesian product of varieties with diagonal $\hat{\mathbb{Z}}$–action.*

PROOF. The endofunctors σ_n of $\mathcal{C}^{\hat{\mathbb{Z}}}$ map an object (X,α) to $(X, \alpha\circ\sigma_n)$ and a pair of $\hat{\mathbb{Z}}$-equivariant embeddings

$$(Y, \alpha|_Y) \hookrightarrow (X,\alpha) \hookleftarrow (X \smallsetminus Y, \alpha|_{X\smallsetminus Y})$$

to the pair of embedding

$$(Y, \alpha|_Y \circ \sigma_n) \hookrightarrow (X, \alpha\circ\sigma_n) \hookleftarrow (X \smallsetminus Y, \alpha|_{X\smallsetminus Y} \circ \sigma_n).$$

This determines the functor σ_n on both objects and morphisms of $\mathcal{C}^{\hat{\mathbb{Z}}}$. The compatibility with the monoidal structure comes from the compatibility with Cartesian products $\sigma_n(X,\alpha)\times\sigma_n(X',\alpha') = (X\times X', (\alpha\times\alpha')\circ\Delta\circ\sigma_n) = \sigma_n((X,\alpha)\times(X',\alpha'))$. The group homomorphisms $\tilde{\rho}_n$ of $K_0^{\hat{\mathbb{Z}}}(\mathcal{V})$ are also induced by endofunctors of $\mathcal{C}^{\hat{\mathbb{Z}}}$, which map objects by $\tilde{\rho}_n(X,\alpha) = (X \times Z_n, \Phi_n(\alpha))$ and pairs of $\hat{\mathbb{Z}}$–equivariant embeddings

$$(Y, \alpha|_Y) \hookrightarrow (X,\alpha) \hookleftarrow (X \smallsetminus Y, \alpha|_{X\smallsetminus Y})$$

to pair of embedding

$$(Y \times Z_n, \Phi_n(\alpha|_Y)) \hookrightarrow (X \times Z_n, \Phi_n(\alpha)) \hookleftarrow ((X \smallsetminus Y) \times Z_n, \Phi_n(\alpha|_{X\smallsetminus Y}))$$

where $\Phi_n(\alpha|_Y) = \Phi_n(\alpha)|_Y$ and $\Phi_n(\alpha|_{X\smallsetminus Y}) = \Phi_n(\alpha)|_{X\smallsetminus Y}$. The functors $\tilde{\rho}_n$, however, are not compatible with the monoidal structure, and this reflects the fact that they only induce group homomorphisms on $K_0^{\hat{\mathbb{Z}}}(\mathcal{V})$ rather than ring homomorphisms. ☐

One can obtain a similar argument working with subtractive Waldhausen categories as in [**Ca15**] in place of assemblers as in [**Za17a**].

4.7. The Kontsevich–Tschinkel Burnside ring. In a similar way, instead of working with the Grothendieck ring $K_0^{\hat{\mathbb{Z}}}(\mathcal{V})$, we can consider the refinement of the Grothendieck ring constructed in [**KoTsch17**]. We discuss here briefly how to adapt the previous construction to this case.

In [**KoTsch17**] a refinement of the Grothendieck ring of varieties is introduced, which is based on birational equivalence. More precisely, for $\mathbb{K}$ a field of characteristic zero, the Burnside semiring $\mathrm{Burn}_+(\mathbb{K})$ is defined as the set of equivalence classes of smooth $\mathbb{K}$-varieties under the $\mathbb{K}$-birational equivalence relation, with addition and multiplication are given by disjoint union and product over $\mathbb{K}$ (Definition 2 of [**KoTsch17**]). The Burnside ring $\mathrm{Burn}(\mathbb{K})$ is the Grothendieck ring generated by the semiring $\mathrm{Burn}_+(\mathbb{K})$. Equivalently, the Burnside ring $\mathrm{Burn}(\mathbb{K})$ is generated by isomorphism classes $[X]$ of smooth varieties over $\mathbb{K}$ with the equivalence relation $[X] = [U]$ for $U \hookrightarrow X$ an open embedding with dense image.

To construct an assembler and an associated spectrum that recovers the Burnside ring Burn($\mathbb{K}$) as its zeroth homotopy group, we proceed again as in [**Za17a**].

LEMMA 4.3. *Let $\mathcal{C}_{\rm Burn}$ be the category with non–initial objects given by the smooth $\mathbb{K}$-varieties X and morphisms given by the open embeddings $U \hookrightarrow X$ with dense image. Consider the Grothendieck topology which is generated by the open dense embeddings $U \hookrightarrow X$. The category $\mathcal{C}_{\rm Burn}$ is an assembler and the associated spectrum $K(\mathcal{C}_{\rm Burn})$ has $\pi_0 K(\mathcal{C}_{\rm Burn}) = {\rm Burn}(\mathbb{K})$.*

PROOF. The initial object is the empty scheme. If X is irreducible, a disjoint covering family consists of a single dense open set $U \hookrightarrow X$ and the common refinement of two disjoint covering families $U_1 \hookrightarrow X$ and $U_2 \hookrightarrow X$ is the dense open set $U_1 \cap U_2 \hookrightarrow X$. Morphisms are monomorphisms given by compositions of open dense embeddings. This shows that the category $\mathcal{C}_{\rm Burn}$ is an assembler. As an abelian group, $\pi_0 K(\mathcal{C}_{\rm Burn})$ is generated by the objects of $\mathcal{C}_{\rm Burn}$ with relations $[X] = \sum_i [X_i]$ for $\{f_i : X_i \to X\}$ a finite disjoint covering family. In this case this means identifying $[X] = [U]$ for any dense open embedding $U \hookrightarrow X$, which is the equivalence relation of Burn($\mathbb{K}$). □

It is shown in [**KoTsch17**] that the Burnside ring Burn($\mathbb{K}$) with the grading given by the transcendence degree, maps surjectively to the associated graded object ${\rm gr}\, K_0(\mathcal{V}_{\mathbb{K}})$ with respect to the filtration of $K_0(\mathcal{V}_{\mathbb{K}})$ by dimension

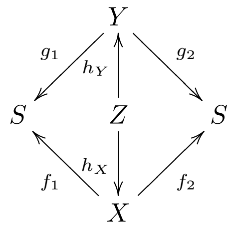
$$ {\rm Burn}(\mathbb{K}) \to {\rm gr}\, K_0(\mathcal{V}_{\mathbb{K}}). \tag{4.2} $$

As we did in the case of the Grothendieck ring, we can also consider an equivariant version of the Kontsevich–Tschinkel Burnside ring Burn($\mathbb{K}$) with respect to the group $\hat{\mathbb{Z}}$, see sec. 5 of [**KoTsch17**]. The corresponding assembler and spectrum are obtained as a modification of the case discussed above. The following statement can be proved by arguments as in Lemma 4.3 and Lemma 4.1.

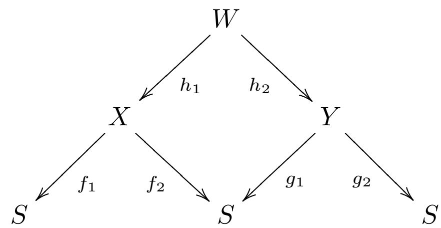
LEMMA 4.4. *Let ${\rm Burn}^{\hat{\mathbb{Z}}}(\mathbb{K})$ be generated by equivalence classes of smooth $\mathbb{K}$-varieties with a good $\hat{\mathbb{Z}}$-action with respect to the equivalence relation $[X] = [U]$ for $U \hookrightarrow X$ a $\hat{\mathbb{Z}}$-equivariant dense open embedding. The category $\mathcal{C}^{\hat{\mathbb{Z}}}_{\rm Burn}$ with objects the smooth $\mathbb{K}$–varieties X with a good $\hat{\mathbb{Z}}$–action and and morphisms the $\hat{\mathbb{Z}}$–equivariant dense open embeddings $U \hookrightarrow X$ is an assembler with $\pi_0 K(\mathcal{C}^{\hat{\mathbb{Z}}}_{\rm Burn}) = {\rm Burn}^{\hat{\mathbb{Z}}}(\mathbb{K})$.*

We refer to $K(\mathcal{C}^{\hat{\mathbb{Z}}}_{\rm Burn})$ as the $\hat{\mathbb{Z}}$–equivariant Burnside spectrum.

The notion of an *epimorphic assembler with a sink* was introduced in Section 4 of [**Za17a**]. It denotes an assembler $\mathcal{C}$ with a sink object S such that ${\rm Hom}(X, S) \neq \emptyset$ for all other objects $X \in \mathcal{C}$, and with the properties that all morphisms $f : X \to Y$ in $\mathcal{C}$ with X non-initial are epimorphisms with the set $\{f : X \to Y\}$ a covering family, and for $X, Y \neq \emptyset$ no two morphisms $X \to Z$ and $Y \to Z$ are disjoint. There is a group $G_{\mathcal{C}}$ associated to epimorphic assembler with a sink, with elements the equivalence classes of pairs of morphisms $f_1, f_2 : X \to S$ from a non-initial object to the sink, where the equivalence $[f_1, f_2 : X \to S] = [g_1, g_2 : Y \to S]$ is determined by the existence of an object Z and maps $h_X : Z \to X$ and $h_Y : Z \to Y$ such that the diagram commutes



and with the composition given by any (equivalent) completion to a commutative diagram of the form



It is shown in Theorem 4.8 of [**Za17a**] that any choice of a morphism $f_X : X \to S$ from each object of $\mathcal{C}$ to the sink object S determines a morphism of assemblers $\mathcal{C} \to \mathbb{S}_G$, where $\mathbb{S}_G$ is the assembler with objects $\emptyset$ and $\star$, a non-invertible morphism $\emptyset \to \star$ and invertible morphisms $\mathrm{Aut}(\star) = G$, which has spectrum $K(\mathbb{S}_G) = \Sigma^\infty_+ BG$ (see Example 3.2 of [**Za17a**]). This morphism of assemblers $\mathcal{C} \to \mathbb{S}_G$ induces an equivalence on K-theory.

Lemma 4.5. *The assembler $\mathcal{C}^{\hat{\mathbb{Z}}}_{\mathrm{Burn}}$ is a coproduct of epimorphic assemblers with sinks*

$$\mathcal{C}^{\hat{\mathbb{Z}}}_{\mathrm{Burn}} = \bigvee_{[X,\alpha]} \mathcal{C}^{\hat{\mathbb{Z}}}_{\mathrm{Burn}}(X, \alpha)$$

where

$$K(\mathcal{C}^{\hat{\mathbb{Z}}}_{\mathrm{Burn}}(X, \alpha)) \simeq \Sigma^\infty_+ B\,\mathrm{Aut}^{\hat{\mathbb{Z}}}(\mathbb{K}(X, \alpha)),$$

with $\mathrm{Aut}^{\hat{\mathbb{Z}}}(\mathbb{K}(X, \alpha))$ *the group of $\hat{\mathbb{Z}}$-equivariant birational automorphisms of X with good $\hat{\mathbb{Z}}$-action α.*

Proof. For an irreducible smooth projective variety X with a good $\hat{\mathbb{Z}}$-action $\alpha : \hat{\mathbb{Z}} \times X \to X$, consider the assembler $\mathcal{C}^{\hat{\mathbb{Z}}}_{\mathrm{Burn}}(X, \alpha)$ with objects $(U, \alpha_U) \hookrightarrow (X, \alpha)$ the $\hat{\mathbb{Z}}$-equivariant dense open embeddings, with α, α_U the compatible good actions of $\hat{\mathbb{Z}}$ on X and U, respectively. Arguing as in Theorem 5.3 of [**Za17a**] for the non-eqiuvariant case, we see that $\mathcal{C}^{\hat{\mathbb{Z}}}_{\mathrm{Burn}}(X)$ satisfies the conditions of an epimorphic assembler with sink. The associated group $G^{\hat{\mathbb{Z}}}_{(X,\alpha)}$ consists of equivalence classes of pairs $f_1, f_2 : (U, \alpha_U) \to (X, \alpha)$, and the f_i are equivariant with respect to these actions. The group $G^{\hat{\mathbb{Z}}}_{(X,\alpha)}$ is therefore given by the group $\mathrm{Aut}^{\hat{\mathbb{Z}}}(\mathbb{K}(X, \alpha))$ of $\hat{\mathbb{Z}}$-equivariant birational automorphisms of the variety with good $\hat{\mathbb{Z}}$ action (X, α). We then have $K(\mathcal{C}^{\hat{\mathbb{Z}}}_{\mathrm{Burn}}(X, \alpha)) \simeq \Sigma^\infty_+ B\,\mathrm{Aut}^{\hat{\mathbb{Z}}}(\mathbb{K}(X, \alpha))$. Moreover, we can identify the assembler $\mathcal{C}^{\hat{\mathbb{Z}}}_{\mathrm{Burn}}$ with the coproduct over equivalence classes $[X, \alpha]$ of the assemblers $\mathcal{C}^{\hat{\mathbb{Z}}}_{\mathrm{Burn}}(X, \alpha)$ as above, since the morphisms of $\mathcal{C}^{\hat{\mathbb{Z}}}_{\mathrm{Burn}}$ between non-initial objects come from morphisms of the $\mathcal{C}^{\hat{\mathbb{Z}}}_{\mathrm{Burn}}(X, \alpha)$ and the objects of $\mathcal{C}^{\hat{\mathbb{Z}}}_{\mathrm{Burn}}$ consist of an initial object $\emptyset$ and the non-initial objects of the $\mathcal{C}^{\hat{\mathbb{Z}}}_{\rm Burn}(X,\alpha)$ for a choice of representatives of the classes $[X,\alpha]$. $\square$

The relation between the Kontsevich–Tschinkel Burnside ring $\mathrm{Burn}^{\hat{\mathbb{Z}}}(\mathbb{K})$ and the equivariant Grothendieck ring $K_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{K}})$ can then be formulated at the level of assemblers and spectra in a form similar to Theorem 5.2 of [**Za17a**], using the same argument, adapted to the equivariant case. Let $\mathcal{C}^{\hat{\mathbb{Z}},(\ell)}_{\mathbb{K}}$ denote the full sub-assembler of the assembler $\mathcal{C}^{\hat{\mathbb{Z}}}_{\mathbb{K}}$ of Lemma 4.1 above, consisting of varieties of dimension at most ℓ with good $\hat{\mathbb{Z}}$-action.

PROPOSITION 4.6. *Let $B^{\hat{\mathbb{Z}}}_n$ denote the set of birational isomorphism classes of varieties of dimension n with good $\hat{\mathbb{Z}}$-action, through $\hat{\mathbb{Z}}$-equivariant birational isomorphisms. The coproduct assembler*

$$\mathcal{C}^{\hat{\mathbb{Z}}}_{\mathrm{Burn},n} := \bigvee_{[X,\alpha]\in B^{\hat{\mathbb{Z}}}_n} \mathcal{C}^{\hat{\mathbb{Z}}}_{\rm Burn}(X,\alpha)$$

satisfies

$$K(\mathcal{C}^{\hat{\mathbb{Z}}}_{\mathrm{Burn},n}) \simeq \bigvee_{[X,\alpha]\in B^{\hat{\mathbb{Z}}}_n} \Sigma^\infty_+ B\,\mathrm{Aut}^{\hat{\mathbb{Z}}}(\mathbb{K}(X,\alpha)) \simeq \mathrm{hocofib}(K(\mathcal{C}^{\hat{\mathbb{Z}},(n-1)}_{\mathbb{K}}) \to K(\mathcal{C}^{\hat{\mathbb{Z}},(n)}_{\mathbb{K}})).$$

4.8. Burnside spectrum and Bost–Connes endomorphisms. The same procedure we used to lift the Bost–Connes maps σ_n and $\tilde{\rho}_n$ to the Grothendieck ring $K_0^{\hat{\mathbb{Z}}}(\mathcal{V})$ and the spectrum $K^{\hat{\mathbb{Z}}}(\mathcal{V})$ can be adapted to lift the same maps to the Kontsevich–Tschinkel Burnside ring $\mathrm{Burn}^{\hat{\mathbb{Z}}}(\mathbb{K})$ and the spectrum $K(\mathcal{C}^{\hat{\mathbb{Z}}}_{\rm Burn})$.

PROPOSITION 4.7. *The maps σ_n and $\tilde{\rho}_n$ of the integral Bost–Connes algebra lift to endofunctors of the assembler category $\mathcal{C}^{\hat{\mathbb{Z}}}_{\rm Burn}$, with the σ_n compatible with the monoidal structure induced by the Cartesian product. These endofunctors induce the corresponding maps σ_n and $\tilde{\rho}_n$ on the Kontsevich–Tschinkel Burnside ring* $\mathrm{Burn}^{\hat{\mathbb{Z}}}(\mathbb{K})$.

PROOF. We argue as in Proposition 4.2. The endofunctors σ_n of $\mathcal{C}^{\hat{\mathbb{Z}}}_{\rm Burn}$ map an object (X,α) to $(X,\alpha\circ\sigma_n)$ and a $\hat{\mathbb{Z}}$-equivariant dense open embedding

$$(U,\alpha|_U)\hookrightarrow (X,\alpha)$$

to the $\hat{\mathbb{Z}}$–equivariant dense open embedding

$$(U,\alpha|_U\circ\sigma_n)\hookrightarrow (X,\alpha\circ\sigma_n).$$

This determines the functor σ_n on both objects and morphisms of $\mathcal{C}^{\hat{\mathbb{Z}}}_{\rm Burn}$. As in Proposition 4.2 one sees the σ_n are compatible with Cartesian products and induce ring homomorphisms of $\mathrm{Burn}^{\hat{\mathbb{Z}}}(\mathbb{K})$. The $\tilde{\rho}_n$ map objects by $\tilde{\rho}_n(X,\alpha) = (X\times Z_n, \Phi_n(\alpha))$ and $\hat{\mathbb{Z}}$-equivariant dense open embeddings $(U,\alpha|_U)\hookrightarrow(X,\alpha)$ by

$$(U\times Z_n,\Phi_n(\alpha|_U))\hookrightarrow (X\times Z_n,\Phi_n(\alpha))$$

with $\Phi_n(\alpha|_U)=\Phi_n(\alpha)|_U$. The $\tilde{\rho}_n$ are not compatible with the monoidal structure and only induce group homomorphism on $\mathrm{Burn}^{\hat{\mathbb{Z}}}(\mathbb{K})$. $\square$

5. Expectation values, motivic measures, and zeta functions

5.1. The Bost–Connes expectation values. In the case of the original Bost–Connes system, one considers representations π of the Bost–Connes algebra (either the integral $\mathcal{A}_{\mathbb{Z}}$ or the rational $\mathcal{A}_{\mathbb{Q}} = \mathbb{Q}[\mathbb{Q}/\mathbb{Z}] \rtimes \mathbb{N}$) on a Hilbert space $\mathcal{H} = \ell^2(\mathbb{N})$ and associates to the algebra and the representation a dynamical system, namely the one–parameter group of automorphism $\sigma : \mathbb{R} \to \mathrm{Aut}\,(\mathcal{A})$ of the C^*-algebra generated by $\mathcal{A}_{\mathbb{Q}}$, seen as an algebra of bounded operators on $\mathcal{H}$. The time evolution satisfies the covariance condition

$$\pi(\sigma_t(a)) = e^{itH}\pi(a)e^{-itH},$$

where H is an (unbounded) linear operator on $\mathcal{H}$, the Hamiltonian of the system. In the Bost–Connes case the time evolution is determined by $\sigma_t(\mu_n) = n^{it}\mu_n$ and $\sigma_t(x) = x$ for $x \in \mathbb{Q}[\mathbb{Q}/\mathbb{Z}]$. The Hamiltonian acts on the standard orthonormal basis of $\ell^2(\mathbb{N})$ as $H\epsilon_n = \log(n)\,\epsilon_n$ and the partition function $Z(\beta) = \mathrm{Tr}(e^{-\beta H})$ is the Riemann zeta function, [**BoCo95**]. For any element $a \in \mathcal{A}_{\mathbb{Q}}$ the expectation value with respect to the Bost–Connes dynamics is then given by

$$\langle a\rangle_\beta = \zeta(\beta)^{-1}\,\mathrm{Tr}(\pi(a)e^{-\beta H}) = \zeta(\beta)^{-1}\sum_{n\in\mathbb{N}}\langle\epsilon_n, \pi(a)e^{-\beta H}\epsilon_n\rangle. \tag{5.1}$$

We can similarly construct Bost–Connes expectation values associated to the non-commutative ring $\mathbf{K}_0^{\hat{\mathbb{Z}}}(\mathcal{V})$ defined in Section 3.4.

5.2. The equivariant Euler characteristic. As we discussed above, the equivariant Euler characteristic $\chi : K_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{Q}}) \to K_0^{\hat{\mathbb{Z}}}(\mathbb{Q})$ induces a ring homomorphism $\mathbf{K}_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{Q}}) \to \mathcal{A}_{\mathbb{Z}}$ where $\mathcal{A}_{\mathbb{Z}}$ is the integral Bost–Connes algebra. After tensoring with $\mathbb{Q}$, one obtains a morphism of crossed product algebras $\mathbf{K}_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{Q}}) \otimes \mathbb{Q} = K_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{Q}})_{\mathbb{Q}} \rtimes \mathbb{N} \to \mathcal{A}_{\mathbb{Q}} = \mathbb{Q}[\mathbb{Q}/\mathbb{Z}] \rtimes \mathbb{N}$, with $K_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{Q}})_{\mathbb{Q}} = K_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{Q}}) \otimes \mathbb{Q}$.

PROPOSITION 5.1. *Let π be a representation of the Bost–Connes algebra $\mathcal{A}_{\mathbb{Q}}$ on the Hilbert space $\mathcal{H} = \ell^2(\mathbb{N})$ with $\pi(\mu_n)\epsilon_m = \epsilon_{nm}$ and $\pi(e(r))\epsilon_n = \zeta_r^n\epsilon_n$ for $r \to \zeta_r$ an embedding of $\mathbb{Q}/\mathbb{Z}$ as the group of roots of unity in $\mathbb{C}^*$. Then π determines a one–parameter family of group homomorphism $\varphi_\beta : K_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{Q}}) \to \mathbb{C}$, with $\beta \in \mathbb{R}_+^*$, such that for all $[X,\alpha] \in K_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{Q}})$ the product $\zeta(\beta)\cdot\langle[X,\alpha]\rangle_\beta$, with $\zeta(\beta)$ the Riemann zeta funciton, is a $\mathbb{Z}$-combination of values at roots of unity of the polylogarithm function $\mathrm{Li}_\beta(x)$.*

PROOF. For the generators $a = e(r)$ of $\mathbb{Z}[\mathbb{Q}/\mathbb{Z}]$ the expectation value (5.1) is a polylogarithm function evaluated at a root of unity normalized by the Riemann zeta function,

$$\langle e(r)\rangle_\beta = \zeta(\beta)^{-1}\sum_{n\geq 1}\zeta_r^n\, n^{-\beta} = \frac{\mathrm{Li}_\beta(\zeta_r)}{\zeta(\beta)},$$

where $\pi(e(r))\epsilon_n = \zeta_r^n\epsilon_n$ with $r \mapsto \zeta_r$ an embedding of $\mathbb{Q}/\mathbb{Z}$ as the roots of unity in $\mathbb{C}^*$. Given a representation π of the Bost–Connes algebra, we compose the equivariant Euler characteristic $\mathbf{K}_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{Q}}) \otimes \mathbb{Q} = K_0^{\hat{\mathbb{Z}}}(\mathcal{V}_{\mathbb{Q}})_{\mathbb{Q}} \rtimes \mathbb{N} \to \mathcal{A}_{\mathbb{Q}} = \mathbb{Q}[\mathbb{Q}/\mathbb{Z}] \rtimes \mathbb{N}$ with the Bost–Connes expectation value $\varphi(X,\alpha) = \langle\chi(X,\alpha)\rangle_\beta$. ∎

5.3. Expectation values of motivic measures. Other examples can be constructed using other motivic measures. For example, one can consider the mixed Hodge motivic measure $h : K_0(\mathcal{V}_{\mathbb{C}}) \to K_0(HS)$ with $h(X) = \sum_r (-1)^r [H^r_c(X,\mathbb{Q})] \in K_0(HS)$. This is a refinement of the Hodge–Deligne polynomial motivic measure $P : K_0(\mathcal{V}_{\mathbb{C}}) \to \mathbb{Z}[u,v,u^{-1},v^{-1}]$ with $P(X,u,v) = \sum_{p,q} \dim H^{p,q} u^p v^q$. In the case of complex varieties with a good $\hat{\mathbb{Z}}$-action that factors through a finite quotient $\mathbb{Z}/n\mathbb{Z}$, the graded pieces $H^{p,q}$ are $\hat{\mathbb{Z}}$-modules. The equivariant Hodge–Deligne polynomial is then defined as the polynomial in $\mathbb{Z}[\mathbb{Q}/\mathbb{Z}][u,v]$ given by $P^{\hat{\mathbb{Z}}}((X,\alpha),u,v) = \sum_{p,q} E^{p,q}(X,\alpha) u^q v^q$ with $E^{p,q}(X,\alpha) = \sum_k (-1)^k H^{p,q}(H^k_c(X,\mathbb{C}))$, with the $\hat{\mathbb{Z}}$-module structure determined by the actoin α. The equivariant weight polynomial is given by $W^{\hat{\mathbb{Z}}}(X,w) = P^{\hat{\mathbb{Z}}}(X,w,w)$ while evaluation at $w=1$ recovers the equivariant Euler characteristic. The associated expectation values are then of the form

$$\varphi_\beta(X,\alpha) = \sum_{p,q} \langle E^{p,q}(X,\alpha)\rangle_\beta \; u^q v^q.$$

5.4. Zeta functions and assemblers. Passing from the level of Grothendieck rings to assemblers, spectra, and K-theory, as in [**Za17a**]–[**Za17c**], also provides possible methods for lifting the zeta functions at the level of K-theory. One approach, currently being developed [**Za18**], directly uses assemblers and the construction of a map of assemblers between the assemblers underlying the Grothendieck ring (and its equivariant version as discussed above) and an assembler of almost-finite-G-sets, by mapping a variety X to the almost-finite set $X(\bar{\mathbb{K}})$. Another approach to the lifting of zeta functions was developed in [**CaWoZa17**], using étale cohomology and SW-categories. Zeta functions and the lifts of the Bost–Connes system to assemblers and spectra are further developed in the second part of this work, [**LieManMar19**].

Following the approach being developed in [**Za18**], one can show that the equivariant Euler characteristic

$$\chi^{\hat{\mathbb{Z}}} : K_0^{\hat{\mathbb{Z}}}(\mathcal{V}) \to \mathbb{Z}[\mathbb{Q}/\mathbb{Z}]$$

discussed above in Sections 3.1–3.2 lifts to a map of assembler by considering, as in Section 3.2, the morphism

$$\chi^G : K_0^G(\mathcal{V}) \to A(G) \to R(G)$$

with $A(G)$ the Burnside ring, for a finite group G, with the equivariant Euler characteristics defined as in [**Gu-Za17**] as mapping $\chi^G(X) = \sum_k [X_k]$ with $[X_k]$ the classes in $A(G)$ of the k-skeleta. In the case of $\hat{\mathbb{Z}}$ one considers the completion $\hat{A}(\hat{\mathbb{Z}}) = \varprojlim A(\mathbb{Z}/n\mathbb{Z})$ as discussed in Section 3.2, where the complete Burnstein ring $\hat{A}(\hat{\mathbb{Z}})$ is seen as the Grothendieck ring of almost-finite $\hat{\mathbb{Z}}$-sets [**DrSi88**]. According to [**Za18**], there is a construction of an assembler of almost-finite-G-sets, which we denote by $\mathcal{AF}^G$. The equivariant Euler characteristic

$$\chi^G : K_0^G(\mathcal{V}) \to A(G)$$

then lifts to a morphism of assemblers

$$\chi^G : \mathcal{C}^G \to \mathcal{AF}^G$$

since the assignment of X to the union of the X_k as G-sets and G-equivariant embeddings $Y \hookrightarrow X$ and $X \smallsetminus Y \hookrightarrow X$ to the corresponding maps of the skeleta as

G-sets maps disjointness morphisms in the assembler $\mathcal{C}^G$ to disjoint morphisms in the assembler $\mathcal{AF}^G$. In particular, the equivariant Euler characteristic

$$\chi^{\hat{\mathbb{Z}}} : K_0^{\hat{\mathbb{Z}}}(\mathcal{V}) \to \hat{A}(\hat{\mathbb{Z}})$$

can be lifted to a morphism of assemblers

$$\chi^{\hat{\mathbb{Z}}} : \mathcal{C}^{\hat{\mathbb{Z}}} \to \mathcal{AF}^{\hat{\mathbb{Z}}}.$$

6. Dynamical $\mathbb{F}_1$-structures and the Bost–Connes algebra

6.1. The spectrum as Euler characteristic. The point of view we adopt here is similar to [**EbGu-Za17**]. We consider the Grothendieck ring $K_0^{\mathbb{Z}}(\mathcal{V}_{\mathbb{C}})$ of pairs (X, f) of a complex quasi-projective variety X with an automorphism $f : X \to X$, such that f_* in homology is quasi-unipotent. The addition is given by disjoint union and the product by the Cartesian product.

The quasi-unipotent condition ensures that the spectrum of the induced action $f_* : H_*(X, \mathbb{Z}) \to H_*(X, \mathbb{Z})$ is contained in the set of roots of unity. We can then consider the spectrum of f_* as an Euler characteristic.

Lemma 6.1. *The spectrum of the induced map on homology determines a ring homomorphism*

$$\sigma : K_0^{\mathbb{Z}}(\mathcal{V}_{\mathbb{C}}) \to \mathbb{Z}[\mathbb{Q}/\mathbb{Z}]. \tag{6.1}$$

Proof. To a pair (X, f) we associate the spectrum of the map $f_* : H_*(X, \mathbb{Z}) \to H_*(X, \mathbb{Z})$, seen as a subset $\Sigma(f_*) \subset \mathbb{Q}/\mathbb{Z}$ of roots of unity counted with integer multiplicities. Thus, we have $\sigma(X, f) = \sum_{\lambda \in \Sigma(f_*)} m_\lambda \, \lambda$. The spectrum of a tensor product is given by the set of products of eigenvalues of the two matrices, hence the compatibility with the ring structure of $K_0^{\mathbb{Z}}(\mathcal{V}_{\mathbb{C}})$. $\square$

Under suitable assumptions on the induced map on $H^*(X, \mathbb{C})$ and its Hodge decomposition, one can also consider other kinds of motivic measures associated to the spectrum Σ, for example generalizations of the Hodge–Deligne polynomial, see [**EbGu-Za17**].

6.2. Lifting the Bost–Connes algebra to dynamical $\mathbb{F}_1$-structures. Let Z_n be a zero-dimensional variety with $\#Z_n(\mathbb{C}) = n$. Then, for a given $(X, f) \in K_0^{\mathbb{Z}}(\mathcal{V}_{\mathbb{C}})$, the *Verschiebung* pair $(X \times Z_n, \Phi_n(f))$ consists of the variety $X \times Z_n$ with the automorphism $\Phi_n(f)(x, a_i) = (x, a_{i+1})$ for $i = 1, \dots, n-1$ and $\Phi_n(f)(x, a_n) = (f(x), a_1)$.

Lemma 6.2. *The induced map in homology $\Phi_n(f)_* : H_*(X \times Z_n, \mathbb{Z}) \to H_*(X \times Z_n, \mathbb{Z})$ is the Verschiebung map.*

Proof. We have $H_k(Z_n, \mathbb{Z}) = \mathbb{Z}^n$ for $k = 0$ and zero otherwise, hence we can identify $H_*(X \times Z_n, \mathbb{Z}) \simeq H_*(X, \mathbb{Z})^{\oplus n}$. Then the action $\Phi_n(f)(x, a_i) = (x, a_{i+1})$ for $i = 1, \dots, n-1$ and $\Phi_n(f)(x, a_n) = (f(x), a_1)$ induces the action $\Phi_n(f)_* = V(f_*)$ in homology. $\square$

Proposition 6.3. *The maps $\sigma_n(X, f) = (X, f^n)$ and $\tilde{\rho}_n(X, f) = (X \times Z_n, \Phi_n(f))$ lift the maps σ_n and $\tilde{\rho}_n$ of the integral Bost–Connes algebra to the Grothendieck ring $K_0^{\mathbb{Z}}(\mathcal{V}_{\mathbb{C}})$, compatibly with the spectrum Euler characteristic $\sigma : K_0^{\mathbb{Z}}(\mathcal{V}_{\mathbb{C}}) \to \mathbb{Z}[\mathbb{Q}/\mathbb{Z}]$.*

PROOF. The argument is analogous to the case of $\hat{\mathbb{Z}}$-actions analyzed in the previous section. Because of the relations between Frobenius $F_n(f) = f^n$ and Verschiebung $V_n(f)$ we have

$$\sigma_n \circ \tilde{\rho}_n(X,f) = \sigma_n(X \times Z_n, \Phi_n(f)) = (X \times Z_n, \Phi_n(f)^n) = (X \times Z_n, f \times 1) = (X,f)^{\oplus n}$$

$$\tilde{\rho}_n \circ \sigma_n(X,f) = \tilde{\rho}_n(X,f^n) = (X \times Z_n, \Phi_n(f^n)) = (X,f) \times (Z_n, \gamma),$$

with $\gamma = \Phi_n(1) : a_i \mapsto a_{i+1}$ and $a_n \mapsto a_1$, where as before we used the relation $V_n(F_n(a)b) = aV_n(b)$, which gives $\Phi_n(f^n) = f\Phi_n(1)$. Under the spectrum Euler characteristic map $\sigma : K_0^{\mathbb{Z}}(\mathcal{V}_{\mathbb{C}}) \to \mathbb{Z}[\mathbb{Q}/\mathbb{Z}]$ we then see that we have commutative diagrams

$$\begin{array}{ccc} K_0^{\mathbb{Z}}(\mathcal{V}_{\mathbb{C}}) & \xrightarrow{\sigma} & \mathbb{Z}[\mathbb{Q}/\mathbb{Z}] \\ \downarrow{\scriptstyle \sigma_n} & & \downarrow{\scriptstyle \sigma_n} \\ K_0^{\mathbb{Z}}(\mathcal{V}_{\mathbb{C}}) & \xrightarrow{\sigma} & \mathbb{Z}[\mathbb{Q}/\mathbb{Z}] \end{array}$$

and similarly for the $\tilde{\rho}_n$, where σ_n are ring homomorphism and $\tilde{\rho}_n$ are group homomorphisms. □

Thus, we can consider a non-commutative version of the Grothendieck ring $\mathbf{K}_0^{\mathbb{Z}}(\mathcal{V}_{\mathbb{C}})$.

DEFINITION 6.4. Let $\mathbf{K}_0^{\mathbb{Z}}(\mathcal{V}_{\mathbb{C}})$ be the non-commutative ring generated by $K_0^{\mathbb{Z}}(\mathcal{V}_{\mathbb{C}})$ together with generators $\tilde{\mu}_n$ and μ_n^* satisfying the relations (3.1) for all $n, m \in \mathbb{N}$, and (3.2) for all $x = (X,f) \in K_0^{\mathbb{Z}}(\mathcal{V}_{\mathbb{C}})$ and all $n \in \mathbb{N}$.

Consider then the algebra $K_0^{\mathbb{Z}}(\mathcal{V}_{\mathbb{C}})_{\mathbb{Q}} = K_0^{\mathbb{Z}}(\mathcal{V}_{\mathbb{C}}) \otimes_{\mathbb{Z}} \mathbb{Q}$. As in the $\hat{\mathbb{Z}}$–equivariant case analyzed in the previous section, the maps σ_n and $\tilde{\rho}_n$ induce endomorphisms σ_n and ρ_n of $K_0^{\mathbb{Z}}(\mathcal{V}_{\mathbb{C}})_{\mathbb{Q}}$, which determine a non-commutative semigroup crossed product algebra. The spectrum Euler characteristic (6.1) extends to an algebra homomorphism to the rational Bost–Connes algebra.

PROPOSITION 6.5. *The algebra $\mathbf{K}_0^{\mathbb{Z}}(\mathcal{V}_{\mathbb{C}}) \otimes_{\mathbb{Z}} \mathbb{Q}$ is isomorphic to a semigroup crossed product algebra $K_0^{\mathbb{Z}}(\mathcal{V}_{\mathbb{C}})_{\mathbb{Q}} \rtimes \mathbb{N}$ with the semigroup action given by $x \mapsto n^{-1}\tilde{\rho}_n(x)$. The spectrum Euler characteristic* (6.1) *extends to an algebra homomorphism $\sigma : \mathbf{K}_0^{\mathbb{Z}}(\mathcal{V}_{\mathbb{C}}) \otimes_{\mathbb{Z}} \mathbb{Q} \to \mathcal{A}_{\mathbb{Q}}$, where $\mathcal{A}_{\mathbb{Q}} = \mathbb{Q}[\mathbb{Q}/\mathbb{Z}] \rtimes \mathbb{N}$ is the rational Bost–Connes algebra.*

PROOF. The algebra $\mathbf{K}_0^{\mathbb{Z}}(\mathcal{V}_{\mathbb{C}}) \otimes_{\mathbb{Z}} \mathbb{Q}$ is generated by the elements of $K_0^{\mathbb{Z}}(\mathcal{V}_{\mathbb{C}})_{\mathbb{Q}}$ and additional generators $\mu_n = n^{-1}\tilde{\mu}_n$ and μ_n^*, which satisfy the relations

$$\mu_n^*\mu_n = 1, \quad \mu_{nm} = \mu_n\mu_m, \quad \mu_{nm}^* = \mu_n^*\mu_m^*, \quad \forall n, m \in \mathbb{N},$$

$$\mu_n\mu_m^* = \mu_m^*\mu_n \text{ if } (n,m) = 1$$

$$\mu_n x \mu_n^* = \rho_n(x) \quad \text{with} \quad \rho_n(x) = \frac{1}{n}\tilde{\rho}_n(x),$$

with $\sigma_n\rho_n(x) = x$, for all $x = (X,f) \in K_0^{\mathbb{Z}}(\mathcal{V}_{\mathbb{C}})$. The semigroup action in the crossed product algebra $K_0^{\mathbb{Z}}(\mathcal{V}_{\mathbb{C}})_{\mathbb{Q}} \rtimes \mathbb{N}$ is given by $x \mapsto \rho_n(x) = \mu_n x \mu_n^*$, hence one obtains an identification of these two algebras. The morphism $\sigma : \mathbf{K}_0^{\mathbb{Z}}(\mathcal{V}_{\mathbb{C}}) \otimes_{\mathbb{Z}} \mathbb{Q} \to \mathcal{A}_{\mathbb{Q}}$ is the map given by the spectrum Euler characteristic on elements of $K_0^{\mathbb{Z}}(\mathcal{V}_{\mathbb{C}})$, extended to $\mathbb{Q}$-coefficients, and it maps $\chi(\mu_n) = \mu_n$ and $\chi(\mu_n^*) = \mu_n^*$. By Proposition 6.3, it determines a homomorphism of crossed–product algebras. □

6.3. Assembler and Bost–Connes endofuntors. First we consider an assembler category $\mathcal{C}_{\mathbb{C}}^{\mathbb{Z}}$ associated to the ring $K_0^{\mathbb{Z}}(\mathcal{V}_{\mathbb{C}})$ and the associated Γ–space and spectrum $K^{\mathbb{Z}}(\mathcal{V}_{\mathbb{C}})$ with $\pi_0 K^{\mathbb{Z}}(\mathcal{V}_{\mathbb{C}}) = K_0^{\mathbb{Z}}(\mathcal{V}_{\mathbb{C}})$, then we show that the maps σ_n and $\tilde{\rho}_n$ define endofunctors of the assembler $\mathcal{C}_{\mathbb{C}}^{\mathbb{Z}}$, hence they induce maps of spectra and induced map of the homotopy groups that recover the Bost–Connes map on π_0.

PROPOSITION 6.6. *Let $\mathcal{C}_{\mathbb{C}}^{\mathbb{Z}}$ be the following category. Its objects are the pairs (X, f) of a complex quasi–projective variety X with an automorphism $f : X \to X$, such that the induced map f_* in homology is quasi–unipotent. Its morphisms $\varphi : (Y, h) \hookrightarrow (X, f)$ are given by embeddings $Y \hookrightarrow X$ of components preserved by the map f and $h = f|_Y$. This is an assembler category, and the associated spectrum $K^{\mathbb{Z}}(\mathcal{V}_{\mathbb{C}}) := K(\mathcal{C}_{\mathbb{C}}^{\mathbb{Z}})$ has $\pi_0 K(\mathcal{C}_{\mathbb{C}}^{\mathbb{Z}}) = K_0^{\mathbb{Z}}(\mathcal{V}_{\mathbb{C}})$. The maps σ_n and $\tilde{\rho}_n$ on $K_0^{\mathbb{Z}}(\mathcal{V}_{\mathbb{C}})$ lift to endofunctors of the assembler $\mathcal{C}_{\mathbb{C}}^{\mathbb{Z}}$, in which σ_n also compatible with the monoidal structure.*

PROOF. The argument is similar to the $\hat{\mathbb{Z}}$-equivariant case we discussed before. In the category $\mathcal{C}_{\mathbb{C}}^{\mathbb{Z}}$ the Grothendieck topology is generated by the covering families $\{(X_1, f|_{X_1}) \hookrightarrow (X, f), (X_2, f|_{X_2}) \hookrightarrow (X, f)\}$ with $X = X_1 \sqcup X_2$ and the X_i are preserves by the map $f : X \to X$. The empty X is the initial object. The finite disjoint covering families are given by embeddings $\varphi_i : (X_i, f|_{X_i}) \hookrightarrow (X, f)$, where the X_i are unions of components preserved by the map, $f|_{X_i} = f \circ \varphi_i$. Any two finite disjoint families have a common refinement since the category has pullbacks, [**Za17a**] and morphisms are compositions of embeddings hence monomorphisms. The abelian group structure on $\pi_0 K(\mathcal{C}_{\mathbb{C}}^{\mathbb{Z}})$ is determined by the relation $(X, f) = (X_1, f|_{X_1}) + (X_1, f|_{X_2})$ for each decomposition $X = X_1 \sqcup X_2$ that is preserved by the map $f : X \to X$. The product is determined by the symmetric monoidal structure induced by the Cartesian product. Thus, we obtain the ring $K_0^{\mathbb{Z}}(\mathcal{V}_{\mathbb{C}})$. The endofunctors σ_n map objects by $\sigma_n(X, f) = (X, f^n)$ and maps pairs of embeddings with $X = X_1 \sqcup X_2$

$$\{(X_1, f|_{X_1}) \hookrightarrow (X, f) \hookleftarrow (X_2, f|_{X_2})\}$$

to pairs of embeddings

$$\{(X_1, f^n|_{X_1}) \hookrightarrow (X, f^n) \hookleftarrow (X_2, f^n|_{X_2})\}$$

These functors are compatible with Cartesian products, hence with the monoidal structure. The endofunctore $\tilde{\rho}_n$ act on objects as $\tilde{\rho}_n(X, f) = (X \times Z_n, \Phi_n(f))$ and map a pair of embeddings as above to the pair

$$\{(X_1 \times Z_n, \Phi_n(f)|_{X_1}) \hookrightarrow (X \times Z_n, \Phi_n(f)) \hookleftarrow (X_2 \times Z_n, \Phi_n(f)|_{X_2})\},$$

where $\Phi_n(f)|_{X_i} = \Phi_n(f|_{X_i})$. The functors $\tilde{\rho}_n$ are not compatible with the monoidal structure hence they induce group homomorphisms of $\pi_0 K(\mathcal{C}_{\mathbb{C}}^{\mathbb{Z}})$. $\square$

Note that, unlike the $\hat{\mathbb{Z}}$-equivariant cases considered in the previous sections, the spectrum $K(\mathcal{C}_{\mathbb{C}}^{\mathbb{Z}})$ is not so interesting topologically, since in the assembler we are only using decompositions into connected components. The reason for wanting only this type of scissor-congruence relations in $K_0^{\mathbb{Z}}(\mathcal{V}_{\mathbb{C}})$ is the spectrum Euler characteristic $\sigma : K_0^{\mathbb{Z}}(\mathcal{V}_{\mathbb{C}}) \to \mathbb{Z}[\mathbb{Q}/\mathbb{Z}]$, which should map a splitting $X = X_1 \sqcup X_2$ compatible with $f : X \to X$ to a corresponding splitting $H_*(X, \mathbb{Z}) = H_*(X_1, \mathbb{Z}) \oplus H_*(X_2, \mathbb{Z})$ with quasi-unipotent maps $f_*|_{X_i}$, so that the spectrum as an element of $\mathbb{Z}[\mathbb{Q}/\mathbb{Z}]$ satisfies $\sigma(X, f) = \sigma(X_1, f_1) + \sigma(X_2, f_2)$.

Acknowledgment. We thank Inna Zakharevich for several useful comments.

References

[At61] M. F. Atiyah, *Characters and cohomology of finite groups*, Inst. Hautes Études Sci. Publ. Math. **9** (1961), 23–64. MR0148722

[Bo11a] James Borger, *The basic geometry of Witt vectors, I: The affine case*, Algebra Number Theory **5** (2011), no. 2, 231–285, DOI 10.2140/ant.2011.5.231. MR2833791

[Bo11b] James Borger, *The basic geometry of Witt vectors. II: Spaces*, Math. Ann. **351** (2011), no. 4, 877–933, DOI 10.1007/s00208-010-0608-1. MR2854117

[Bor14] Lev A. Borisov, *The class of the affine line is a zero divisor in the Grothendieck ring*, J. Algebraic Geom. **27** (2018), no. 2, 203–209, DOI 10.1090/jag/701. MR3764275

[BoCo95] J.-B. Bost and A. Connes, *Hecke algebras, type III factors and phase transitions with spontaneous symmetry breaking in number theory*, Selecta Math. (N.S.) **1** (1995), no. 3, 411–457, DOI 10.1007/BF01589495. MR1366621

[BousFr78] A. K. Bousfield and E. M. Friedlander, *Homotopy theory of Γ-spaces, spectra, and bisimplicial sets*, Geometric applications of homotopy theory (Proc. Conf., Evanston, Ill., 1977), II, Lecture Notes in Math., vol. 658, Springer, Berlin, 1978, pp. 80–130. MR513569

[Ca15] Jonathan A. Campbell, *The K-theory spectrum of varieties*, Trans. Amer. Math. Soc. **371** (2019), no. 11, 7845–7884, DOI 10.1090/tran/7648. MR3955537

[CaWoZa17] J. Campbell, J. Wolfson, I. Zakharevich, *Derived ℓ-adic zeta functions*, arXiv:1703.09855 (2017).

[CC10] Alain Connes and Caterina Consani, *Schemes over $\mathbb{F}_1$ and zeta functions*, Compos. Math. **146** (2010), no. 6, 1383–1415, DOI 10.1112/S0010437X09004692. MR2735370

[CC14] Alain Connes and Caterina Consani, *On the arithmetic of the BC-system*, J. Noncommut. Geom. **8** (2014), no. 3, 873–945, DOI 10.4171/JNCG/173. MR3261604

[CC16] Alain Connes and Caterina Consani, *Absolute algebra and Segal's Γ-rings: au dessous de $\overline{\operatorname{Spec}(\mathbb{Z})}$*, J. Number Theory **162** (2016), 518–551, DOI 10.1016/j.jnt.2015.12.002. MR3448278

[CCM09] Alain Connes, Caterina Consani, and Matilde Marcolli, *Fun with $\mathbb{F}_1$*, J. Number Theory **129** (2009), no. 6, 1532–1561, DOI 10.1016/j.jnt.2008.08.007. MR2521492

[De02] P. Deligne, *Catégories tensorielles* (French, with English summary), Mosc. Math. J. **2** (2002), no. 2, 227–248, DOI 10.17323/1609-4514-2002-2-2-227-248. Dedicated to Yuri I. Manin on the occasion of his 65th birthday. MR1944506

[DrSi88] Andreas W. M. Dress and Christian Siebeneicher, *The Burnside ring of profinite groups and the Witt vector construction*, Adv. in Math. **70** (1988), no. 1, 87–132, DOI 10.1016/0001-8708(88)90052-7. MR947758

[EbGu-Za17] Wolfgang Ebeling and Sabir M. Gusein-Zade, *Higher-order spectra, equivariant Hodge-Deligne polynomials, and Macdonald-type equations*, Singularities and computer algebra, Springer, Cham, 2017, pp. 97–108. MR3675723

[EKMM97] A. D. Elmendorf, I. Kriz, M. A. Mandell, and J. P. May, *Rings, modules, and algebras in stable homotopy theory*, Mathematical Surveys and Monographs, vol. 47, American Mathematical Society, Providence, RI, 1997. With an appendix by M. Cole. MR1417719

[FrSh81] John Franks and Michael Shub, *The existence of Morse-Smale diffeomorphisms*, Topology **20** (1981), no. 3, 273–290, DOI 10.1016/0040-9383(81)90003-3. MR608601

[Gr81] Daniel R. Grayson, *SK_1 of an interesting principal ideal domain*, J. Pure Appl. Algebra **20** (1981), no. 2, 157–163, DOI 10.1016/0022-4049(81)90089-X. MR601681

[Gu-Za17] S. M. Guseĭn-Zade, *Equivariant analogues of the Euler characteristic and Macdonald type formulas* (Russian, with Russian summary), Uspekhi Mat. Nauk **72** (2017), no. 1(433), 3–36, DOI 10.4213/rm9748; English transl., Russian Math. Surveys **72** (2017), no. 1, 1–32. MR3608029

[HSS00] Mark Hovey, Brooke Shipley, and Jeff Smith, *Symmetric spectra*, J. Amer. Math. Soc. **13** (2000), no. 1, 149–208, DOI 10.1090/S0894-0347-99-00320-3. MR1695653

[JMOS16] Seokho Jin, Wenjun Ma, Ken Ono, and Kannan Soundararajan, *Riemann hypothesis for period polynomials of modular forms*, Proc. Natl. Acad. Sci. USA **113** (2016), no. 10, 2603–2608, DOI 10.1073/pnas.1600569113. MR3482847

[KapSmi] M. Kapranov, A. Smirnov, *Cohomology determinants and reciprocity laws: number field case*, Unpublished manuscript.

[KoTsch17] M. Kontsevich, Yu. Tschinkel. *Specialisation of birational types*. arXiv:1708.05699 (2017).

[La09] Tyler Lawson, *Commutative* Γ*-rings do not model all commutative ring spectra*, Homology Homotopy Appl. **11** (2009), no. 2, 189–194. MR2559641

[Le81] H. W. Lenstra Jr., *Grothendieck groups of abelian group rings*, J. Pure Appl. Algebra **20** (1981), no. 2, 173–193, DOI 10.1016/0022-4049(81)90091-8. MR601683

[LeBr17] Lieven Le Bruyn, *Linear recursive sequences and* $\mathtt{Spec}(\mathbb{Z})$ *over* $\mathbb{F}_1$, Comm. Algebra **45** (2017), no. 7, 3150–3158, DOI 10.1080/00927872.2016.1236116. MR3594588

[LeBr19] L. Le Bruyn, *Motivic measures and* F_1*-geometries*, arXiv:1901.10243 (2019).

[LieManMar19] J.F. Lieber, Yu.I. Manin, M. Marcolli, *Bost-Connes systems and* F_1*-structures in Grothendieck rings, spectra, and Nori motives*, arXiv:1901.00020 (2019).

[Lo99] Eduard Looijenga, *Motivic measures*, Astérisque **276** (2002), 267–297. Séminaire Bourbaki, Vol. 1999/2000. MR1886763

[LoLo11] Javier López Peña and Oliver Lorscheid, *Torified varieties and their geometries over* $\mathbb{F}_1$, Math. Z. **267** (2011), no. 3-4, 605–643, DOI 10.1007/s00209-009-0638-0. MR2776051

[Ly99] Manos Lydakis, *Smash products and* Γ*-spaces*, Math. Proc. Cambridge Philos. Soc. **126** (1999), no. 2, 311–328, DOI 10.1017/S0305004198003260. MR1670245

[MacL71] Saunders MacLane, *Categories for the working mathematician*, Springer-Verlag, New York-Berlin, 1971. Graduate Texts in Mathematics, Vol. 5. MR0354798

[Ma95] Yuri Manin, *Lectures on zeta functions and motives (according to Deninger and Kurokawa)*, Astérisque **228** (1995), 4, 121–163. Columbia University Number Theory Seminar (New York, 1992). MR1330931

[Ma10] Yuri I. Manin, *Cyclotomy and analytic geometry over* $\mathbb{F}_1$, Quanta of maths, Clay Math. Proc., vol. 11, Amer. Math. Soc., Providence, RI, 2010, pp. 385–408. MR2732059

[Ma16] Yu. I. Manin, *Local zeta factors and geometries under* Spec $\mathbf{Z}$, Izv. Ross. Akad. Nauk Ser. Mat. **80** (2016), no. 4, 123–130, DOI 10.4213/im8392. Reprinted in Izv. Math. **80** (2016), no. 4, 751–758. MR3535360

[ManMar16] Yuri I. Manin and Matilde Marcolli, *Moduli operad over* $\mathbb{F}_1$, Absolute arithmetic and $\mathbb{F}_1$-geometry, Eur. Math. Soc., Zürich, 2016, pp. 331–361. MR3701904

[Mar09] Matilde Marcolli, *Cyclotomy and endomotives*, p-Adic Numbers Ultrametric Anal. Appl. **1** (2009), no. 3, 217–263, DOI 10.1134/S2070046609030042. MR2566053

[MaRe17] Matilde Marcolli and Zhi Ren, *q-deformations of statistical mechanical systems and motives over finite fields*, p-Adic Numbers Ultrametric Anal. Appl. **9** (2017), no. 3, 204–227, DOI 10.1134/S2070046617030049. MR3682249

[MaTa17] Matilde Marcolli and Gonçalo Tabuada, *Bost-Connes systems, categorification, quantum statistical mechanics, and Weil numbers*, J. Noncommut. Geom. **11** (2017), no. 1, 1–49, DOI 10.4171/JNCG/11-1-1. MR3626556

[Mart16] Nicolas Martin, *The class of the affine line is a zero divisor in the Grothendieck ring: an improvement* (English, with English and French summaries), C. R. Math. Acad. Sci. Paris **354** (2016), no. 9, 936–939, DOI 10.1016/j.crma.2016.05.016. MR3535349

[Mazo12] Volodymyr Mazorchuk, *Lectures on algebraic categorification*, QGM Master Class Series, European Mathematical Society (EMS), Zürich, 2012. MR2918217

[OnRoSp16] Ken Ono, Larry Rolen, and Florian Sprung, *Zeta-polynomials for modular form periods*, Adv. Math. **306** (2017), 328–343, DOI 10.1016/j.aim.2016.10.004. MR3581305

[Schw99] Stefan Schwede, *Stable homotopical algebra and Γ-spaces*, Math. Proc. Cambridge Philos. Soc. **126** (1999), no. 2, 329–356, DOI 10.1017/S0305004198003272. MR1670249

[Schw12] S. Schwede, *Symmetric spectra*, preprint 2012, `https://www.math.uni-bonn.de/~schwede/SymSpec-v3.pdf`

[Se74] Graeme Segal, *Categories and cohomology theories*, Topology **13** (1974), 293–312, DOI 10.1016/0040-9383(74)90022-6. MR0353298

[ShSu75] M. Shub and D. Sullivan, *Homology theory and dynamical systems*, Topology **14** (1975), 109–132, DOI 10.1016/0040-9383(75)90022-1. MR0400306

[Sou04] Christophe Soulé, *Les variétés sur le corps à un élément* (French, with English and Russian summaries), Mosc. Math. J. **4** (2004), no. 1, 217–244, 312, DOI 10.17323/1609-4514-2004-4-1-217-244. MR2074990

[ToVa09] Bertrand Toën and Michel Vaquié, *Au-dessous de* $\operatorname{Spec}\mathbb{Z}$ (French, with English and French summaries), J. K-Theory **3** (2009), no. 3, 437–500, DOI 10.1017/is008004027jkt048. MR2507727

[Ve73] Jean-Louis Verdier, *Caractéristique d'Euler-Poincaré* (French), Bull. Soc. Math. France **101** (1973), 441–445. MR0356109

[Za17a] Inna Zakharevich, *The K-theory of assemblers*, Adv. Math. **304** (2017), 1176–1218, DOI 10.1016/j.aim.2016.08.045. MR3558230

[Za17b] Inna Zakharevich, *On K_1 of an assembler*, J. Pure Appl. Algebra **221** (2017), no. 7, 1867–1898, DOI 10.1016/j.jpaa.2016.12.028. MR3614978

[Za17c] Inna Zakharevich, *The annihilator of the Lefschetz motive*, Duke Math. J. **166** (2017), no. 11, 1989–2022, DOI 10.1215/00127094-0000016X. MR3694563

[Za18] I. Zakharevich, private communication, July 2018.

Max Planck Institute for Mathematics, Bonn, Germany
Email address: `manin@mpim-bonn.mpg.de`

California Institute of Technology, Pasadena, USA
Current address: University of Toronto, Toronto, Canada
Current address: Perimeter Institute for Theoretical Physics, Waterloo, Canada
Email address: `matilde@caltech.edu`

Contemporary Mathematics
Volume **744**, 2020
https://doi.org/10.1090/conm/744/14979

Dynamical zeta functions of Reidemeister type and representations spaces

Alexander Fel'shtyn and Malwina Zietek

Dedicated to the memory of Sergiy Kolyada

Abstract. In this paper we continue to study the Reidemeister zeta function. We prove Pólya – Carlson dichotomy between rationality and a natural boundary for analytic behavior of the Reidemeister zeta function for a large class of automorphisms of Abelian groups. We also study dynamical representation theory zeta functions counting numbers of fixed irreducible representations for iterations of an endomorphism. The rationality and functional equation for these zeta functions are proven for several classes of groups. We find a connection between these zeta functions and the Reidemeister torsions of the corresponding mapping tori. We also establish the connection between the Reidemeister zeta function and dynamical representation theory zeta functions under restriction of endomorphism to a subgroup and to a quotient group.

1. Introduction

Let G be a countable discrete group and $\phi : G \to G$ an endomorphism. Two elements $\alpha, \beta \in G$ are said to be *ϕ-conjugate* or *twisted conjugate,* iff there exists $g \in G$ with $\beta = g\alpha\phi(g^{-1})$. We shall write $\{x\}_\phi$ for the *ϕ-conjugacy* or *twisted conjugacy* class of the element $x \in G$. The number of ϕ-conjugacy classes is called the *Reidemeister number* of an endomorphism ϕ and is denoted by $R(\phi)$. If ϕ is the identity map then the ϕ-conjugacy classes are the usual conjugacy classes in the group G. Taking a dynamical point of view, we consider the iterates of ϕ, and we may define [**11**] a Reidemeister zeta function of ϕ as a power series:

$$R_\phi(z) = \exp\left(\sum_{n=1}^{\infty} \frac{R(\phi^n)}{n} z^n\right).$$

2010 *Mathematics Subject Classification.* Primary 37C25; 37C30; 22D10; Secondary 20E45; 54H20; 55M20.

Key words and phrases. Twisted conjugacy class, Reidemeister number, Reidemeister zeta function, unitary dual.

The work is funded by the Narodowe Centrum Nauki of Poland (NCN) (grant No. 2016/23/G/ST1/04280 (Beethoven 2)).

Whenever we mention the Reidemeister zeta function $R_\phi(z)$, we shall assume that it is well-defined and so $R(\phi^n) < \infty$ for all $n > 0$. The following problem was investigated [**13**]: for which groups and endomorphisms is the Reidemeister zeta function a rational function? Is this zeta function an algebraic function?

When a Reidemeister zeta function is a rational function the infinite sequence of coefficients of the corresponding power series is closely interconnected, and is given by the finite set of zeros and poles of the zeta function. In [**11–14, 21**], the rationality of the Reidemeister zeta function $R_\phi(z)$ was proven in the following cases: the group is finitely generated and an endomorphism is eventually commutative; the group is finite; the group is a direct sum of a finite group and a finitely generated free Abelian group; the group is finitely generated, nilpotent and torsion free. Recently, the rationality and functional equation for the Reidemeister zeta function were proven for endomorphisms of fundamental groups of infra-nilmanifolds [**5**] and for endomorphisms of fundamental groups of infra-solvmanifolds of type (R) [**15**].

In this paper we continue to study the Reidemeister zeta function. We prove Pólya – Carlson dichotomy between rationality and a natural boundary for analytic behavior of Reidemeister zeta function for a large class of automorphisms of Abelian groups.

We continue to study dynamical representation theory zeta functions (see [**18**]) counting numbers of fixed irreducible unitary representations for iterates of an endomorphism. The rationality and functional equation for these zeta functions are proven for several classes of groups. We find a connection between these zeta functions and the Reidemeister torsions of the corresponding mapping tori.

We establish the connection between Reidemeister zeta function and dynamical representation theory zeta functions under restriction of endomorphism to a subgroup and to a quotient group.

Our method is to identify the Reidemeister numbers with the number of fixed points of the induced map $\hat\phi$ (respectively, its iterations) of an appropriate subspace of the unitary dual $\widehat{G}$, when $R(\varphi) < \infty$. This method is called the twisted Burnside–Frobenius theory (TBFT), because in the case of a finite group and identity automorphism we arrive to the classical Burnside–Frobenius theorem on enumerating of (usual) conjugacy classes via irreducible unitary representations.

Let us present the contents of the paper in more details.

In Section 2 the rationality and functional equation for dynamical representation theory zeta functions are proven for endomorphisms of finitely generated Abelian groups; endomorphisms of finitely generated torsion free nilpotent groups; endomorphisms of groups with finite ϕ-irreducible subspaces of corresponding unitary dual spaces and for automorphisms of crystallographic groups with diagonal holonomy Z_2. For a periodic automorphism of a group we have proved a product formula for dynamical representation theory zeta functions which implies that these zeta functions are radicals of rational functions.

In Section 3 we investigate the rationality of these zeta functions and the connections between Reidemeister zeta function and dynamical representation theory zeta functions under restriction of endomorphism to a subgroup and to a quotient group. We also prove the Gauss congruences for the Reidemeister numbers of iterations of endomorphism for a group with polycyclic quotient group.

In Section 4 we presents results in support of a Pólya – Carlson dichotomy between rationality and a natural boundary for analytic behavior of Reidemeister zeta function for a large class of automorphisms of Abelian groups.

2. Dynamical zeta functions and representations spaces

Suppose, ϕ is an endomorphism of a discrete group G. Denote by $\widehat{G}$ the *unitary dual* of G, i.e. the space of equivalence classes of unitary irreducible representations of G, equipped with the *hull-kernel* topology, denote by $\widehat{G}_f$ the subspace of the unitary dual formed by irreducible finite-dimensional representations, and by $\widehat{G}_{ff}$ the subspace of $\widehat{G}_f$ formed by *finite* representations, i.e. representations that factorize through a finite group. Generally the correspondence $\widehat{\phi}: \rho \mapsto \rho\circ\phi$ does not define a dynamical system (an action of the semigroup of positive integers) on the unitary dual $\widehat{G}$ or its finite-dimensional part $\widehat{G}_f$, or finite part $\widehat{G}_{ff}$, because in contrast with the authomorphism case, the representation $\rho\circ\phi$ may be reducible, so it is only possible to decompose $\rho\circ\phi$ into irreducible components and we obtain a sort of multivalued map $\widehat{\phi}$.

Nevertheless we can consider representations ρ such that $\rho\sim\rho\circ\phi$ and proceed as follows.

Definition 2.1. A *representation theory Reidemeister number* $RT(\phi)$ *is defined* [**18**] as the number of all $[\rho]\in\widehat{G}$ such that $\rho\sim\rho\circ\phi$. Taking $[\rho]\in\widehat{G}_f$ (respectively $[\rho]\in\widehat{G}_{ff}$) we obtain $RT^f(\phi)$ (respectively $RT^{ff}(\phi)$). Evidently $RT(\phi)\geqslant RT^f(\phi)\geqslant RT^{ff}(\phi)$.

In analogy with the Reidemeister zeta function and other similar objects we have defined in [**18**] jointly with E.Troitsky following dynamical representation zeta functions

$$RT_\phi(z)=\exp\left(\sum_{n=1}^\infty \frac{RT(\phi^n)}{n}z^n\right),$$
$$RT^f_\phi(z)=\exp\left(\sum_{n=1}^\infty \frac{RT^f(\phi^n)}{n}z^n\right),$$
$$RT^{ff}_\phi(z)=\exp\left(\sum_{n=1}^\infty \frac{RT^{ff}(\phi^n)}{n}z^n\right),$$

when numbers $RT(\phi^n)$ (resp, $RT^f(\phi^n)$, or $RT^{ff}(\phi^n)$) are all finite.

The importance of these numbers is justified by the following dynamical interpretation. In [**17**] the following "dynamical part" of the dual space, where $\widehat{\phi}$ and all its iterations $\widehat{\phi}^n$ define a dynamical system, was defined.

Definition 2.2. Following [**17**] a class $[\rho]$ is called a $\widehat{\phi}$-**f**-point, if $\rho\sim\rho\circ\phi$ (so, these are the points under consideration in the Definition 2.1).

Definition 2.3. Following [**17**] an element $[\rho]\in\widehat{G}$ (respectively, in $\widehat{G}_f$ or $\widehat{G}_{ff}$) is called *ϕ-irreducible* if $\rho\circ\phi^n$ is irreducible for any $n=0,1,2,\dots$.

Denote the corresponding subspaces of $\widehat{G}$ (resp., $\widehat{G}_f$ or $\widehat{G}_{ff}$) by $\widehat{G}^\phi$ (resp., $\widehat{G}^\phi_f$ or $\widehat{G}^\phi_{ff}$).

LEMMA 2.4 (Lemma 2.4 in [**17**]). *Suppose the representations ρ and $\rho \circ \phi^n$ are equivalent for some $n \geq 1$. Then $[\rho] \in \widehat{G}^\phi$.*

COROLLARY 2.5 (Corollary 2.5 in [**17**]). *Generally, there is no dynamical system defined by $\widehat{\phi}$ on $\widehat{G}$ (resp., $\widehat{G}_f$, or $\widehat{G}_{ff}$). We have only the well-defined notion of a $\widehat{\phi}^n$-**f**-point.*

*A well-defined dynamical system exists on $\widehat{G}^\phi$ (resp., $\widehat{G}^\phi_f$, or $\widehat{G}^\phi_{ff}$). Its n-periodic points are exactly $\widehat{\phi}^n$-**f**-points.*

We refer to [**17**] for proofs and details.

Once we have identified the coefficients of representation theory zeta functions with the numbers of periodic points of a dynamical system, the standard argument with the Möbius inversion formula (see e.g. [**12**, p. 104], [**17**]) gives the following statement.

THEOREM 2.6. *(Theorem 2.7 of [**18**]) Suppose , $RT(\phi^n) < \infty$ for any n. Then we have the following Gauss congruences for representation theory Reidemeister numbers:*

$$\sum_{d|n} \mu(d) \cdot RT(\phi^{n/d}) \equiv 0 \mod n$$

for any n.

A similar statement is true for $RT^f(\phi^n)$ and $RT^{ff}(\phi^n)$.

Here the above *Möbius function* is defined as

$$\mu(d) = \begin{cases} 1 & \text{if } d = 1, \\ (-1)^k & \text{if } d \text{ is a product of } k \text{ distinct primes,} \\ 0 & \text{if } d \text{ is not square} - \text{free.} \end{cases}$$

DEFINITION 2.7. Following [**17**] we say that TBFT (resp., TBFT$_f$, TBFT$_{ff}$) takes place for an endomorphism $\phi : G \to G$ and its iterations, if $R(\phi^n) < \infty$ and $R(\phi^n)$ coincides with the number of $\widehat{\phi}^n$-**f**-points in $\widehat{G}$ (resp., in $\widehat{G}_f$, $\widehat{G}_{ff}$) for all $n \in \mathbb{N}$.

Similarly, one can give a definition for a single endomorphism (without iterations).

The following statement follows from the definitions.

PROPOSITION 2.8. *(Proposition 2.8 of [**18**]) Suppose, $\phi : G \to G$ is an endomorphism and $R(\phi) < \infty$. If TBFT (resp., TBFT$_f$) is true for G and ϕ, then $R(\phi) = RT(\phi)$ (resp., $R(\phi) = RT^f(\phi) = RT^{ff}(\phi)$).*

If the suppositions hold for ϕ^n, for any n, then $R_\phi(z) = RT_\phi(z)$ (resp., $R_\phi(z) = RT^f_\phi(z) = RT^{ff}_\phi(z)$).

Denote by $AM^f(\phi^n)$ the number of *isolated* n-periodic points(i.e. *isolated* $\widehat{\phi}^n$-**f**-points) of the dynamical system $(\widehat{\phi})^n$ on $\widehat{G}^\phi_f$.

If these numbers are finite for all powers of ϕ, the corresponding Artin–Masur representation zeta function is defined as

$$AM^f_\phi(z) = \exp\left(\sum_{n=1}^{\infty} \frac{AM^f(\phi^n)}{n} z^n\right).$$

Let $Z(\phi)$ be one of the numbers $RT(\phi)$, $RT^f(\phi)$, $RT^{ff}(\phi)$, $AM^f(\phi)$. Let

$$Z_\phi(z) = \exp\left(\sum_{n=1}^{\infty} \frac{Z(\phi^n)}{n} z^n\right)$$

be one of the zeta functions $AM^f_\phi(z)$, $RT_\phi(z)$, $RT^f_\phi(z)$, $RT^{ff}_\phi(z)$.

THEOREM 2.9. *Let ϕ be a periodic automorphism of least period m of a group G . Then the zeta function $Z_\phi(z)$ is equal to*

$$Z_\phi(z) = \prod_{d|m} \sqrt[d]{(1-z^d)^{-P(d)}},$$

where the product is taken over all divisors d of the period m, and $P(d)$ is the integer

$$P(d) = \sum_{d_1|d} \mu(d_1) Z(\phi^{d|d_1}).$$

PROOF. Since $\phi^m = id$, then $(\widehat{\phi})^m = id$ as well and $Z(\phi^j) = Z(\phi^{m+j})$ for every j. If $(k,m) = 1$, there exist positive integers t and q such that $kt = mq + 1$. So $(\phi^k)^t = \phi^{kt} = \phi^{mq+1} = \phi^{mq}\phi = (\phi^m)^q\phi = \phi$. Consequently, $Z(\phi^k) = Z(\phi)$. The same argument shows that $Z(\phi^d) = Z(\phi^{di})$ if $(i, m/d) = 1$ where d divisor m Using these series of equal numbers we obtain by direct calculation

$$\begin{aligned} Z_\phi(z) &= \exp\left(\sum_{i=1}^{\infty} \frac{Z(\phi^i)}{i} z^i\right) = \exp\left(\sum_{d|m}\sum_{i=1}^{\infty} \frac{P(d)}{d} \cdot \frac{z^{d^i}}{i}\right) \\ &= \exp\left(\sum_{d|m} \frac{P(d)}{d} \cdot \log(1-z^d)\right) = \prod_{d|m} \sqrt[d]{(1-z^d)^{-P(d)}} \end{aligned}$$

where the integers $P(d)$ are calculated recursively by the formula

$$P(d) = Z(\phi^d) - \sum_{d_1|d;\, d_1 \neq d} P(d_1).$$

Moreover, if the last formula is rewritten in the form $Z(\phi^d) = \sum_{d_1|d} P(d_1)$ and one uses the Möbius inversion law for real function in number theory, then

$$P(d) = \sum_{d_1|d} \mu(d_1) \cdot Z(\phi^{d/d_1}),$$

where $\mu(d_1)$ is the Möbius function in number theory. The theorem is proved. □

COROLLARY 2.10. *If in Theorem 2.9 the period m is a prime number, then*

$$Z_\phi(z) = \frac{1}{(1-z)^{Z(\phi)}} \cdot \sqrt[m]{(1-z^m)^{Z(\phi)-Z(\phi^m)}}.$$

THEOREM 2.11. *Let $\phi : G \to G$ be an endomorphism of group G. Suppose that subspaces $\widehat{G}^\phi$, $\widehat{G}^\phi_f$, and $\widehat{G}^\phi_{ff}$ are finite. Then zeta function $Z_\phi(z)$ is a rational function satisfying a functional equation*

$$Z_\phi\left(\frac{1}{z}\right) = (-1)^a z^b Z_\phi(z).$$

In particular we have

$$Z_\phi(z) = \prod_{[\gamma]} \frac{1}{1 - z^{\#[\gamma]}}, \tag{1}$$

where the product is taken over the periodic orbits of the dynamical system $(\widehat{\phi})^n$ *in* $\widehat{G}^\phi$*, resp* $\widehat{G}^\phi_f$*, or* $\widehat{G}^\phi_{ff}$*. In the functional equation the numbers a and b are respectively the number of periodic* $\widehat{\phi}$*-orbits of elements of* $\widehat{G}^\phi$*, resp* $\widehat{G}^\phi_f$*, or* $\widehat{G}^\phi_{ff}$ *and the number of periodic elements of* $\widehat{G}^\phi$*, resp* $\widehat{G}^\phi_f$*, or* $\widehat{G}^\phi_{ff}$*.*

Proof. We shall call an element of $\widehat{G}^\phi$, resp $\widehat{G}^\phi_f$, or $\widehat{G}^\phi_{ff}$ periodic if it is fixed by some iteration of $\widehat{\phi}$. A periodic element γ is fixed by $\widehat{\phi}^n$ iff n is divisible by the cardinality the orbit of γ. We therefore have

$$Z(\phi^n) = \sum_{\substack{\gamma \ periodic \\ \#[\gamma]|n}} 1 = \sum_{\substack{[\gamma] \ such \ that, \\ \#[\gamma]|n}} \#[\gamma].$$

From this follows

$$\begin{aligned} Z_\phi(z) &= \exp\left(\sum_{n=1}^{\infty} \frac{Z(\phi^n)}{n} z^n\right) = \exp\left(\sum_{[\gamma]} \sum_{\substack{n=1 \\ \#[\gamma]|n}}^{\infty} \frac{\#[\gamma]}{n} z^n\right) \\ &= \prod_{[\gamma]} \exp\left(\sum_{n=1}^{\infty} \frac{\#[\gamma]}{\#[\gamma]n} z^{\#[\gamma]n}\right) = \prod_{[\gamma]} \exp\left(\sum_{n=1}^{\infty} \frac{1}{n} z^{\#[\gamma]n}\right) \\ &= \prod_{[\gamma]} \exp\left(-\log\left(1 - z^{\#[\gamma]}\right)\right) = \prod_{[\gamma]} \frac{1}{1 - z^{\#[\gamma]}}. \end{aligned}$$

Moreover

$$\begin{aligned} Z_\phi\left(\frac{1}{z}\right) &= \prod_{[\gamma]} \frac{1}{1 - z^{-\#[\gamma]}} = \prod_{[\gamma]} \frac{z^{\#[\gamma]}}{z^{\#[\gamma]} - 1} = \prod_{[\gamma]} \frac{-z^{\#[\gamma]}}{1 - z^{\#[\gamma]}} \\ &= \prod_{[\gamma]} -z^{\#[\gamma]} Z_\phi(z) = (-1)^{\#\{[\gamma]\}} z^{\sum \#[\gamma]} Z_\phi(z). \end{aligned}$$

□

2.1. Endomorphisms of finitely generated Abelian groups. For a finitely generated Abelian group G we define the finite subgroup G^{finite} to be the subgroup of torsion elements of G. We denote the quotient $G^\infty := G/G^{finite}$. The group G^∞ is torsion free. Since the image of any torsion element by a homomorphism must be a torsion element, the function $\phi : G \to G$ induces maps

$$\phi^{finite} : G^{finite} \longrightarrow G^{finite}, \quad \phi^\infty : G^\infty \longrightarrow G^\infty.$$

If G is abelian, then $\widehat{G} = \widehat{G}_f = \widehat{G}^\varphi = \widehat{G}^\varphi_f$ [**17**].

The Lefschetz zeta function of a discrete dynamical system $\hat{\phi}$ equals:

$$L_{\hat{\phi}}(z) := \exp\left(\sum_{n=1}^{\infty} \frac{L(\hat{\phi}^n)}{n} z^n\right),$$

where

$$L(\hat{\phi}^n) := \sum_{k=0}^{\dim X} (-1)^k \operatorname{Tr}\left[\hat{\phi_{*k}}^n : H_k(\hat{G};Q) \to H_k(\hat{G};Q)\right]$$

is the Lefschetz number of $\hat{\phi}^n$. The Lefschetz zeta function $L_{\hat{\phi}}(z)$ is a rational function of z and is given by the formula:

$$L_{\hat{\phi}}(z) = \prod_{k=0}^{\dim X} \det\left(I - \hat{\phi_{*k}} \cdot z\right)^{(-1)^{k+1}}.$$

THEOREM 2.12. *Let $\phi : G \to G$ be an endomorphism of a finitely generated Abelian group. Then we have*

$$Z(\phi^n) = \mid L(\hat{\phi}^n) \mid, \tag{2}$$

where $L(\hat{\phi}^n)$ is the Lefschetz number of $\hat{\phi}$ thought of as a self-map of the topological space $\hat{G}$. From this it follows that zeta functon $Z_\phi(z)$ is a rational function and is equal to:

$$Z_\phi(z) = L_{\hat{\phi}}(\sigma z)^{(-1)^r}, \tag{3}$$

where $\sigma = (-1)^p$ where p is the number of real eingevalues $\lambda \in \operatorname{Spectr}(\phi^\infty)$ such that $\lambda < -1$ and r is the number of real eingevalues $\lambda \in \operatorname{Spectr}(\phi^\infty)$ such that $\mid \lambda \mid > 1$. If G is a finite abelian group then this reduces to

$$Z(\phi^n) = L(\hat{\phi}^n) \text{ and } Z_\phi(z) = L_{\hat{\phi}}(z).$$

PROOF. If G is finite abelian then $\hat{G}$ is a discrete finite set, so the number of fixed points is equal to the Lefschetz number. This finishes the proof in the case that G is finite.

If G is a finitely generated free Abelian group then the dual of G is a torus whose dimension is equal to the rank of G. The dual of any finitely generated discrete Abelian group is the direct sum of a torus and a finite group.

If G a finitely generated Abelian group it is only necessary to check that the number of fixed points of $\hat{\phi}^n$ is equal to the absolute value of its Lefschetz number. We assume without loss of generality that $n = 1$. We are assuming that $Z(\phi)$ is finite, so the fixed points of $\hat{\phi}$ form a discrete set. We therefore have

$$L(\hat{\phi}) = \sum_{x \in \operatorname{Fix} \hat{\phi}} \operatorname{ind}(\hat{\phi}, x).$$

Since ϕ is a group endomorphism, the zero element $0 \in \hat{G}$ is always fixed. Let x be any fixed point of $\hat{\phi}$. We then have a commutative diagram

$$\begin{array}{ccccc} g & \hat{G} & \overset{\hat{\phi}}{\longrightarrow} & \hat{G} & g \\ \updownarrow & \updownarrow & & \updownarrow & \updownarrow \\ g+x & \hat{G} & \overset{\hat{\phi}}{\longrightarrow} & \hat{G} & g+x \end{array}$$

in which the vertical functions are translations on $\hat{G}$ by x. Since the vertical maps map 0 to x, we deduce that

$$\operatorname{ind}(\hat{\phi}, x) = \operatorname{ind}(\hat{\phi}, 0)$$

and so all fixed points have the same index. It is now sufficient to show that $\mathrm{ind}(\hat{\phi}, 0) = \pm 1$. This follows because the map on the torus

$$\hat{\phi} : \hat{G}_0 \to \hat{G}_0$$

lifts to a linear map of the universal cover, which is in this case the Lie algebra of $\hat{G}$. The index is then the sign of the determinant of the identity map minus this lifted map. This determinant cannot be zero, because $1 - \hat{\phi}$ must have finite kernel by our assumption that the $Z(\phi)$ is finite (if $\det(1 - \hat{\phi}) = 0$ then the kernel of $1 - \hat{\phi}$ is a positive dimensional subgroup of $\hat{G}$, and therefore infinite). So we have $Z(\varphi^n) = \mid L(\hat{\phi}^n) \mid = (-1)^{r+pn} L(\hat{\phi}^n)$ for all n (see also [**12**]).

Then the zeta function

$$Z_\phi(z) = L_{\hat{\phi}}(\sigma z)^{(-1)^r}$$

is rational function as well. □

2.1.1. *Functional equation.* To write down a functional equation for the Reidemeister type zeta functions, we recall the following functional equation for the Lefschetz zeta function:

LEMMA 2.13 ([**19**, Proposition 8], see also [**7**]). *Let M be a closed orientable manifold of dimension m and let $f : M \to M$ be a continuous map of degree d. Then*

$$L_f\left(\frac{\alpha}{dz}\right) = \epsilon\,(-\alpha dz)^{(-1)^m \chi(M)}\, L_f(\alpha z)^{(-1)^m}$$

where $\alpha = \pm 1$ and $\epsilon \in \mathbb{C}$ is a non-zero constant such that if $|d| = 1$ then $\epsilon = \pm 1$.

We obtain:

THEOREM 2.14 (Functional Equation). *Let $\phi : G \to G$ be an endomorphism of a finitely generated Abelian group of the rank ≥ 1. Then the zeta function $Z_\phi(z)$, whenever it is defined, has the following functional equation:*

$$Z_\phi\left(\frac{1}{dz}\right) = Z_\phi(z)^{(-1)^m} \epsilon^{(-1)^r}$$

where d is a degree $\hat{\phi}$, $m = \dim \hat{G}$, ϵ is a constant in $\mathbb{C}^\times$, $\sigma = (-1)^r$, p is the number of real eigenvalues of ϕ^∞ which are > 1 and r is the number of real eingevalues $\lambda \in \mathrm{Spectr}(\phi^\infty)$ such that $\mid \lambda \mid > 1$. If $|d| = 1$ then $\epsilon = \pm 1$.

PROOF. We have $Z_\phi(z) = L_{\hat{\phi}}(\sigma z)^{(-1)^r}$. By Lemma 2.13

$$\begin{aligned} Z_\phi\left(\frac{1}{dz}\right) = L_{\hat{\phi}}\left(\frac{\sigma}{dz}\right)^{(-1)^r} &= \left(\epsilon(-\sigma dz)^{(-1)^m \chi(\hat{G})} L_{\hat{\phi}}(\sigma z)^{(-1)^m}\right)^{(-1)^r} \\ &= Z_\phi(z)^{(-1)^m} \epsilon^{(-1)^r} (-\sigma dz)^{(-1)^{m+r} \chi(\hat{G})}. \end{aligned}$$

On the other hand $\chi(\hat{G}) = 0$ because the dual $\hat{G}$ of any finitely generated discrete Abelian group of the rank ≥ 1 is the direct sum of a torus of dim ≥ 1 and a finite group, i.e. $\hat{G}$ is a union of finitely many tori. This finishes our proof. □

2.2. Endomorphisms of nilpotent groups and crystallographic groups.

THEOREM 2.15. *Let $\phi : G \to G$ be an endomorphism of a finitely generated torsion free nilpotent group G or let ϕ be an automorphism of crystallographic group G with diagonal holonomy Z_2. Then the zeta function $RT_\phi^f(z) = RT_\phi^{ff}(z))$ is rational function.*

PROOF. Any finitely generated torsion free nilpotent group is a supersolvable, hence, polycyclic group. Any crystallographic group with diagonal holonomy Z_2 is a polycyclic-by-finite group. In [**16**,**17**] twisted Burnside-Frobenius theorem($TBFT_f$ and $TBFT_{ff}$) was proven for endomorphisms of polycyclic groups and for automorphisms of polycyclic-by-finite groups. This theorem implies equality of Reidemeister zeta function $R_\phi(z)$ and zeta function $RT_\phi^f(z) = RT_\phi^{ff}(z))$. In [**12**] the rationality of the Reidemeister zeta function $R_\phi(z)$ was proven for endomorphisms of a finitely generated torsion free nilpotent groups and in [**6**] the rationality of $R_\phi(z)$ was proven for automorphisms of crystallographic groups with diagonal holonomy Z_2. This completes the proof.

□

3. Connection with Reidemeister Torsion

Like the Euler characteristic, the Reidemeister torsion is algebraically defined. Roughly speaking, the Euler characteristic is a graded version of the dimension, extending the dimension from a single vector space to a complex of vector spaces. In a similar way, the Reidemeister torsion is a graded version of the absolute value of the determinant of an isomorphism of vector spaces. Let $d^i : C^i \to C^{i+1}$ be a cochain complex C^* of finite dimensional vector spaces over $\mathbb{C}$ with $C^i = 0$ for $i < 0$ and large i. If the cohomology $H^i = 0$ for all i we say that C^* is *acyclic*. If one is given positive densities Δ_i on C^i then the Reidemeister torsion $\tau(C^*, \Delta_i) \in (0, \infty)$ for acyclic C^* is defined as follows:

DEFINITION 3.1. Consider a chain contraction $\delta^i : C^i \to C^{i-1}$, ie. a linear map such that $d \circ \delta + \delta \circ d = id$. Then $d + \delta$ determines a map $(d+\delta)_+ : C^+ := \oplus C^{2i} \to C^- := \oplus C^{2i+1}$ and a map $(d+\delta)_- : C^- \to C^+$. Since the map $(d+\delta)^2 = id + \delta^2$ is unipotent, $(d+\delta)_+$ must be an isomorphism. One defines $\tau(C^*, \Delta_i) :=\mid \det(d+\delta)_+ \mid$ (see [**20**]).

Reidemeister torsion is defined in the following geometric setting. Suppose K is a finite complex and E is a flat, finite dimensional, complex vector bundle with base K. We recall that a flat vector bundle over K is essentially the same thing as a representation of $\pi_1(K)$ when K is connected. If $p \in K$ is a basepoint then one may move the fibre at p in a locally constant way around a loop in K. This defines an action of $\pi_1(K)$ on the fibre E_p of E above p. We call this action the holonomy representation $\rho : \pi \to GL(E_p)$.

Conversely, given a representation $\rho : \pi \to GL(V)$ of π on a finite dimensional complex vector space V, one may define a bundle $E = E_\rho = (\tilde{K} \times V)/\pi$. Here $\tilde{K}$ is the universal cover of K, and π acts on $\tilde{K}$ by covering tranformations and on V by ρ. The holonomy of E_ρ is ρ, so the two constructions give an equivalence of flat bundles and representations of π.

If K is not connected then it is simpler to work with flat bundles. One then defines the holonomy as a representation of the direct sum of π_1 of the components of K. In this way, the equivalence of flat bundles and representations is recovered.

Suppose now that one has on each fibre of E a positive density which is locally constant on K. In terms of ρ_E this assumption just means $\mid \det \rho_E \mid = 1$. Let V denote the fibre of E. Then the cochain complex $C^i(K;E)$ with coefficients in E can be identified with the direct sum of copies of V associated to each i-cell σ of K. The identification is achieved by choosing a basepoint in each component of K and a basepoint from each i-cell. By choosing a flat density on E we obtain a preferred density Δ_i on $C^i(K,E)$. One defines the R-torsion of (K,E) to be $\tau(K;E) = \tau(C^*(K;E), \Delta_i) \in (0,\infty)$.

3.1. The Reidemeister type zeta functions and the Reidemeister torsion of the mapping Torus. Let $f : X \to X$ be a homeomorphism of a compact polyhedron X. Let $T_f := (X \times I)/(x,0) \sim (f(x),1)$ be the mapping torus of f.

We shall consider the bundle $p : T_f \to S^1$ over the circle S^1. We assume here that E is a flat, complex vector bundle with finite dimensional fibre and base S^1. We form its pullback p^*E over T_f. Note that the vector spaces $H^i(p^{-1}(b), c)$ with $b \in S^1$ form a flat vector bundle over S^1, which we denote H^iF. The integral lattice in $H^i(p^{-1}(b), \mathbb{R})$ determines a flat density by the condition that the covolume of the lattice is 1. We suppose that the bundle $E \otimes H^iF$ is acyclic for all i. Under these conditions D. Fried [**20**] has shown that the bundle p^*E is acyclic, and we have

$$\tau(T_f; p^*E) = \prod_i \tau(S^1; E \otimes H^iF)^{(-1)^i}. \tag{4}$$

Let g be the preferred generator of the group $\pi_1(S^1)$ and let $A = \rho(g)$ where $\rho : \pi_1(S^1) \to GL(V)$. Then the holonomy around g of the bundle $E \otimes H^iF$ is $A \otimes f_i^*$. Since $\tau(E) =\mid \det(I - A) \mid$ it follows from (16) that

$$\tau(T_f; p^*E) = \prod_i \mid \det(I - A \otimes f_i^*) \mid^{(-1)^i}. \tag{5}$$

We now consider the special case in which E is one-dimensional, so A is just a complex scalar λ of modulus one. Then in terms of the rational function $L_f(z)$ we have [**20**]:

$$\tau(T_f; p^*E) = \prod_i \mid \det(I - \lambda . f_i^*) \mid^{(-1)^i} =\mid L_f(\lambda) \mid^{-1}. \tag{6}$$

From this formula and Theorem 2.12 we have

THEOREM 3.2. *Let $\phi : G \to G$ be an automorphism of a finitely generated abelian group G. If G is infinite then one has*

$$\tau\left(T_{\hat{\phi}}; p^*E\right) =\mid L_{\hat{\phi}}(\lambda) \mid^{-1} =\mid Z_\phi(\sigma\lambda) \mid^{(-1)^{r+1}},$$

and if G is finite one has

$$\tau\left(T_{\hat{\phi}}; p^*E\right) =\mid L_{\hat{\phi}}(\lambda) \mid^{-1} =\mid Z_\phi(\lambda) \mid^{-1}.$$

where λ is the holonomy of the one-dimensional flat complex bundle E over S^1, r and σ are the constants described in Theorem 2.12 .

3.2. Examples.

Example 3.3. Let Γ be a locally compact group. The following statements are equivalent(see [**2**]): i) Γ has Kazhdan's Property (T); (ii) 1_Γ is isolated in $\widehat{\Gamma}$; (iii) every finite dimensional irreducible unitary representation of Γ is isolated in $\widehat{\Gamma}$; (iv) some finite dimensional irreducible unitary representation of Γ is isolated in $\widehat{\Gamma}$.

This implies immediately that for an endomorphism of a locally compact group Γ with Kazhdan's Property (T) the following zeta functions coincide: $RT_\phi^f(z) = AM_\phi^f(z)$.

Now let us present some examples of Theorem 2.11 for discrete groups with extreme properties. Suppose, an infinite discrete group G has a finite number of conjugacy classes. Such examples can be found in [**16**].

Example 3.4. For the Osin group (see [**26**]) there is only trivial(1-dimensional) finite-dimensional representation. Indeed, the Osin group is an infinite finitely generated group G with exactly two conjugacy classes. All nontrivial elements of this group G are conjugate. So, the group G is simple, i.e. G has no nontrivial normal subgroup. This implies that group G is not residually finite (by definition of residually finite group). Hence, it is not linear (by Mal'cev theorem) and has no finite-dimensional irreducible unitary representations with trivial kernel. Hence, by simplicity of G, it has no finite-dimensional irreducible unitary representation with nontrivial kernel, except for the trivial one. Let us remark that the Osin group is non-amenable, contains the free group in two generators F_2, and has exponential growth.

Let $\phi : G \to G$ be any endomorphism of Osin group G. Thus, we have the following: $RT^f(\phi^n) = RT^{ff}(\phi^n) = 1$ for all n. This implies that for any endomorphism of Osin group G zeta functions

$$RT_\phi^f(z) = RT_\phi^{ff}(z) = \frac{1}{1-z}$$

are rational.

Example 3.5. For large enough prime numbers p, the first examples of finitely generated infinite periodic groups with exactly p conjugacy classes were constructed by Ivanov as limits of hyperbolic groups. The Ivanov group G is an infinite periodic 2-generator group, in contrast to the Osin group, which is torsion free. The Ivanov group G is also a simple group see [**16**]. The discussion can be completed in the same way as in the case of the Osin group.

Example 3.6. G. Higman, B. H. Neumann, and H. Neumann proved that any locally infinite countable group G can be embedded into a countable group G^* in which all elements except the unit element are conjugate to each other (see [**30**]). The discussion above related to the Osin group remains valid for G^* groups.

4. Reduction to subgroups and quotient groups

4.1. Reduction to subgroups. The following lemma is useful for calculating Reidemeister numbers and zeta functions. It will also be used in the proofs of the theorems of this chapter.

LEMMA 4.1. *Let $\phi : G \to G$ be any endomorphism of any group G, and let H be a subgroup of G with the properties*

$$\phi(H) \subset H$$

$$\forall x \in G \; \exists n \in \mathbb{N} \text{ such that } \phi^n(x) \in H.$$

Then

$$R(\phi) = R(\phi_H),$$

where $\phi_H : H \to H$ is the restriction of ϕ to H. If all the numbers $R(\phi^n)$ are finite then

$$R_\phi(z) = R_{\phi_H}(z).$$

PROOF. Let $x \in G$. Then there is an n such that $\phi^n(x) \in H$. It is known that x is ϕ-conjugate to $\phi^n(x)$ (see Lemma 7 [**13**]). This means that the ϕ-conjugacy class $\{x\}_\phi$ of x has non-empty intersection with H.

Now suppose that $x, y \in H$ are ϕ-conjugate, ie. there is a $g \in G$ such that

$$gx = y\phi(g).$$

We shall show that x and y are ϕ_H-conjugate, ie. we can find a $g \in H$ with the above property. First let n be large enough that $\phi^n(g) \in H$. Then applying ϕ^n to the above equation we obtain

$$\phi^n(g)\phi^n(x) = \phi^n(y)\phi^{n+1}(g).$$

This shows that $\phi^n(x)$ and $\phi^n(y)$ are ϕ_H-conjugate. On the other hand, one knows by Lemma 7 that x and $\phi^n(x)$ are ϕ_H-conjugate, and y and $\phi^n(y)$ are ϕ_H conjugate, so x and y must be ϕ_H-conjugate.

We have shown that the intersection with H of a ϕ-conjugacy class in G is a ϕ_H-conjugacy class in H. We therefore have a map

$$\begin{array}{rccc} Rest: & \mathcal{R}(\phi) & \to & \mathcal{R}(\phi_H) \\ & \{x\}_\phi & \mapsto & \{x\}_\phi \cap H \end{array}$$

This clearly has the two-sided inverse

$$\{x\}_{\phi_H} \mapsto \{x\}_\phi.$$

Therefore *Rest* is a bijection and $R(\phi) = R(\phi_H)$. $\square$

Let $Z(\phi)$ be one of the numbers $RT(\phi), RT^f(\phi), RT^{ff}(\phi)$. We shal write $\mathcal{Z}(\phi)$ for one of the corresponding sets $\mathcal{RT}(\phi), \mathcal{RT}^f(\phi), \mathcal{RT}^{ff}(\phi)$ of equivalence classes of irreducible representations.

LEMMA 4.2. *Let $\phi : G \to G$ be any endomorphism of Abelian-by-finite group G, and let H be a subgroup of G with the properties*

$$\phi(H) \subset H$$

$$\forall x \in G \; \exists n \in \mathbb{N} \text{ such that } \phi^n(x) \in H.$$

Then

$$Z(\phi) = Z(\phi_H),$$

where $\phi_H : H \to H$ is the restriction of ϕ to H. If all the numbers $Z(\phi^n)$ are finite then

$$Z_\phi(z) = Z_{\phi_H}(z).$$

Proof. All irreducible representations of Abelian-by-finite group G are finite dimensional.

Let $\rho : G \to U(V)$ be irreducible representation, and suppose that there is a matrix $M \in U(V)$ with

$$\rho \circ \phi = M \cdot \rho \cdot M^{-1},$$

i.e $\rho \in \mathcal{Z}(\phi)$. Suppose that ρ_H, the restriction of ρ to H, is a reducible representation i.e. there is a decomposition $V = V_1 \oplus V_2$ into H-modules. We shall derive a contradiction. We can find $g \in G$ such that $\rho(g)V_1 \notin V_1$. However for sufficiently large n we have $\phi^n(g) \in H$. This shows that $\rho(H)M^nV_1 \notin M^nV_1$. However since H is ϕ-invariant and V_1 is H-invariant, $M^nV_1 = V_1$. Therefore $\rho(H)M^nV_1 \notin V_1$, which gives us a contradiction. Consequently ρ must be an irreducible representation of H on M^nV. However $M^nV = V$, so the representation ρ_H is irreducible. Clearly the class of ρ_H is the same as the class of $\rho_H \circ \phi_H$, i.e $\rho_H \in \mathcal{Z}(\phi_H)$. We thus have a map

$$Rest : \mathcal{Z}(\phi) \to \mathcal{Z}(\phi_H), \rho \to \rho_H.$$

Now let $\rho_H \in \mathcal{Z}(\phi_H)$ be given. Then there is a matrix M such that

$$\rho \circ \phi_H = M \cdot \rho \cdot M^{-1}.$$

If M' is any other such matrix then $M' \cdot M^{-1}$ commutes with $\rho_H(x)$ for all x. It follows that for $g \in \phi^{-n}(H)$ the element

$$M^{-n} \cdot \rho(\phi^n(g)) \cdot M^n$$

is independent of the chosen M, and depends only on ρ, g and n. Now suppose that $\phi^n(g) = h_1 \in H$ and $\phi^m(g) = h_2 \in H, m > n$. Then $\phi^{m-n}(h_1) = h_2$, which implies

$$M^{m-n} \cdot \rho(h_1) \cdot M^{n-m} = \rho(h_2),$$

and therefore

$$M^{-n} \cdot \rho(\phi^n(g)) \cdot M^n = M^{-m} \cdot \rho(\phi^m(g)) \cdot M^m.$$

The above expression is thus independent of M and n, and depends only on ρ and g. We may therefore define for $g \in G$

$$\bar{\rho}(g) = M^{-n} \cdot \rho(\phi^n(g)) \cdot M^n$$

where n is large enough that $\phi^n(g) \in H$. One can easily checks that $\bar{\rho}$ is a representation of G. Since ρ_H is irreducible it follows that $\bar{\rho}$ is irreducible. One sees immediately that the class of $\bar{\rho}$ is the same as the class of $\bar{\rho} \circ \phi$, i.e $\rho \in \mathcal{Z}(\phi)$. Finally we have

$$Rest(\bar{\rho}) = \rho$$

and since any other extension $\tilde{\rho}$ of ρ to G such that $\tilde{\rho} \in \mathcal{Z}(\phi)$ must satisfy

$$\tilde{\rho}(g) = M^{-n} \cdot \rho(\phi^n(g)) \cdot M^n,$$

we have

$$\overline{Rest(\rho)} = \rho.$$

This shows that Rest is a bijection, so $\mathcal{Z}(\phi) = \mathcal{Z}(\phi_H)$. □

Corollary 4.3. *Let $H = \phi^n(G)$. Suppose that all the numbers $R(\phi^k)$ and $Z(\phi^k)$, $k \in N$ are finite. Then*

$$R(\phi) = R(\phi_H), Z(\phi) = Z(\phi_H), R_\phi(z) = R_{\phi_H}(z), Z_\phi(z) = Z_{\phi_H}(z)$$

.

THEOREM 4.4. *Let $\phi : G \to G$ be any endomorphism of Abelian-by-finite group G, and let H be a subgroup of G with the properties*

$$\phi(H) \subset H$$

$$\forall x \in G \, \exists n \in \mathbb{N} \text{ such that } \phi^n(x) \in H.$$

Suppose that all the numbers $R(\phi^k)$ and $Z(\phi^k)$, $k \in N$ are finite. If one of the following conditions is satisfied:

1. *H is a finitely generated abelian group, or*
2. *H is a finite group, or*
3. *H is a crystallographic group with diagonal holonomy $\mathbb{Z}_2$ and ϕ_H is an automorphism,*

then

$$R_\phi(z) = R_{\phi_H}(z) = Z_{\phi_H}(z) = Z_\phi(z)$$

and these zeta functions are rational functions.

PROOF. TBFT (resp., TBFT$_f$, TBFT$_{ff}$) for an endomorphism $\phi : G \to G$ and its iterations were proven in [**13**] for finitely generated abelian groups and for finite groups. Any crystallographic group with diagonal holonomy $\mathbb{Z}_2$ is a polycyclic-by-finite group and it has only finite dimensional irreducible representations. In [**16, 17**] twisted Burnside-Frobenius theorem($TBFT_f$ and $TBFT_{ff}$) was proven for automorphisms of polycyclic-by-finite groups.

These results implie equality of the Reidemeister zeta function $R_{\phi_H}(z)$ and the zeta function $Z_{\phi_H}(z)$. Hence from Lemma 4.1 and Lemma 4.2 it follows that

$$R_\phi(z) = R_{\phi_H}(z) = Z_{\phi_H}(z) = Z_\phi(z).$$

In [**13**] the rationality of the Reidemeister zeta function $R_\phi(z)$ was proven for endomorphisms of finitely generated abelian groups and for finite groups and in [**6**] the rationality of $R_\phi(z)$ was proven for automorphisms of almost-crystallographic groups with diagonal holonomy Z_2. This completes the proof.

□

4.2. Reduction to Injective Endomorphisms on quotient groups. Let G be a group and $\phi : G \to G$ an endomorphism. We shall call an element $x \in G$ nilpotent if there is an $n \in \mathbb{N}$ such that $\phi^n(x) = e$.

Let N be the set of all nilpotent elements of G.

Let $Z(\phi)$ be one of the numbers $RT^f(\phi), RT^{ff}(\phi)$ and $\mathcal{Z}(\phi)$ one of the corresponding sets of equivalence classes of irreducible representations.

THEOREM 4.5. *The set N is a normal subgroup of G. We have $\phi(N) \subset N$ and $\phi^{-1}(N) = N$. Thus ϕ induces an endomorphism $[\phi/N]$ of the quotient group G/N given by $[\phi/N](xN) := \phi(x)N$. The endomorphism $[\phi/N] : G/N \to G/N$ is injective, and we have*

$$R(\phi) = R([\phi/N]), \quad Z(\phi) = Z([\phi/N]).$$

Let the numbers $R(\phi^n)$ and $Z(\phi^n)$ be all finite. Then

$$R_\phi(z) = R_{[\phi/N]}(z), \quad Z_\phi(z) = Z_{[\phi/N]}(z).$$

If the quotient group G/N is polycyclic then one has the following Gauss congruences for Reidemeister numbers:

$$\sum_{d|n} \mu(d) \cdot R(\phi^{n/d}) \equiv 0 \mod n$$

for any n. If one of the following conditions is satisfied:

1. *The quotient group G/N is a finitely generated abelian group, or*
2. *G/N is a finite group, or*
3. *G/N is a finitely generated torsion free nilpotent group, or*
4. *G/N is a crystallographic group with diagonal holonomy $\mathbb{Z}_2$ and $[\phi/N]$ is an automorphism,*

then

$$R_\phi(z) = R_{[\phi/N]}(z) = Z_{[\phi/N]}(z) = Z_\phi(z)$$

and these zeta functions are rational functions.

Proof. (i) Let $x \in N, g \in G$. Then for some $n \in \mathbb{N}$ we have $\phi^n(x) = e$. Therefore $\phi^n(gxg^{-1}) = \phi^n(gg^{-1}) = e$. This shows that $gxg^{-1} \in N$ so N is a normal subgroup of G.

(ii) Let $x \in N$ and choose n such that $\phi^n(x) = e$. Then $\phi^{n-1}(\phi(x)) = e$ so $\phi(x) \in N$. Therefore $\phi(N) \subset N$.

(iii) If $\phi(x) \in N$ then there is an n such that $\phi^n(\phi(x)) = e$. Therefore $\phi^{n+1}(x) = e$ so $x \in N$. This show that $\phi^{-1}(N) \subset N$. The converse inclusion follows from (ii).

(iv) We shal write $\mathcal{R}(\phi)$ for the set of ϕ-conjugacy classes of elements of G. We shall now show that the map $x \to xN$ induces a bijection $\mathcal{R}(\phi) \to \mathcal{R}([\phi/N])$. Suppose $x, y \in G$ are ϕ-conjugate. Then there is a $g \in G$ with $gx = y\phi(g)$. Projecting to the quotient group G/N we have $gnxN = yN\phi(g)N$, so $gNxN = yN[\phi/N](gN)$. This means that xN and yN are $[\phi/N]$-conjugate in G/N. Conversely suppose that xN and yN are $[\phi/N]$-conjugate in G/N. Then there is a $gN \in G/N$ such that $gNxN = yN[\phi/N](gN)$. In other words $gx\phi(g)^{-1}y^{-1}) = e$. Therefore $\phi^n(g)\phi^n(x) = \phi^n(y)\phi^n(\phi(g))$.

This shows that $\phi^n(x)$ and $\phi^n(y)$ are ϕ-conjugate. However x and $\phi^n(x)$ are ϕ-conjugate as are y and $\phi^n(y)$. Therefore x and y are ϕ-conjugate.

(v) We have shown that x and y are ϕ-conjugate iff xN and yN are $[\phi/N]$-conjugate. From this it follows that $x \to xN$ induces a bijection from $\mathcal{R}(\phi)$ to $\mathcal{R}([\phi/N])$. Therefore $R(\phi) = R([\phi/N])$.

(vi) We shall now show that $Z(\phi) = Z([\phi/N])$. Let $\rho \in \mathcal{Z}(\phi)$ and let M be a transformation for which

$$\rho \circ \phi = M \cdot \rho \cdot M^{-1}$$

if $x \in N$ then there is an $n \in \mathbb{N}$ with $\phi^n(x) = e$. Thus N is contained in the kernel of ρ and there is a representation $[\rho/N]$ of G/N given by

$$[\rho/N](gN) := \rho(g).$$

Since $[\rho/N]$ satisfies identity

$$[\rho/N] \circ [\phi/N] = M \cdot [\rho/N] \cdot M^{-1},$$

we have $[\rho/N] \in \mathcal{Z}([\phi/N])$.

(vii) Conversely if $\rho \in \mathcal{Z}([\phi/N])$ we may construct a $\bar{\rho} \in \mathcal{Z}(\phi)$ by

$$\bar{\rho}(x) := \rho(xN).$$

It is clear that $\overline{[\rho/N]} = \rho$ and $\bar{\rho}/N = \rho$.

In [**16**, **17**] the twisted Burnside-Frobenius theorem($TBFT_f$ and $TBFT_{ff}$) was proven for endomorphisms of polycyclic groups and for automorphisms of polycyclic-by-finite groups. Then from (v) and (vi) it follows that

$$R(\phi^n) = R([\phi/N]^n) = Z([\phi/N]^n) = Z(\phi^n).$$

Gauss congruences now follow from Corollary 2.5 and the general theory of congruences for periodic points (cf. [**31**, **33**]). More precisely, let P_d be the number of periodic points of least period d of the dynamical system of Corollary 2.5. Then $R(\phi^n) = Z(\phi^n) = \sum_{d|n} P_d$.

By the Möbius inversion formula,

$$\sum_{d|n} \mu(d) R(\varphi^{n/d}) = P_n \equiv 0 \mod n,$$

since number P_n is always divisible by n, because P_n is exactly n times the number of orbits of cardinality n.

Twisted Burnside-Frobenius theorem($TBFT_f$ and $TBFT_{ff}$) implies also equality of Reidemeister zeta function $R_\phi(z)$ and zeta function $RT_\phi^f(z) = RT_\phi^{ff}(z)$). In [**12**] the rationality of the Reidemeister zeta function $R_\phi(z)$ was proven for endomorphisms of finitely generated abelian groups and for endomorphisms of finitely generated torsion free nilpotent groups, and in [**6**] the rationality of $R_\phi(z)$ was proven for automorphisms of crystallographic groups with diagonal holonomy Z_2. This completes the proof.

□

5. Pólya – Carlson dichotomy for Reidemeister zeta function

In this section we present results in support of a Pólya–Carlson dichotomy between rationality and a natural boundary for the analytic behaviour of the Reidemeister zeta function of Abelian group endomorphisms.

Let $\phi : G \to G$ be a endomorphism of a countable Abelian group G that is a subgroup of $\mathbb{Q}^d$, where $d \geqslant 1$. Let $R = \mathbb{Z}[t]$ be a polynomial ring. Then the Abelian group G naturally carries the structure of a R-module over the ring $R = \mathbb{Z}[t]$ where multiplication by t corresponds to application of the endomorphism: $tg = \phi(g)$ and extending this in a natural way to polynomials. That is, for $g \in G$ and $f = \sum_{n\in\mathbb{Z}} c_n t^n \in R = \mathbb{Z}[t]$ set

$$fg = \sum_{n\in\mathbb{Z}} c_n t^n g = \sum_{n\in\mathbb{Z}} c_n \phi^n(g),$$

where all but finitely many $c_n \in \mathbb{Z}$ are zero. This is a standard procedure for the study of dual automorphisms of compact Abelian groups, see Schmidt [**28**] for an overview.

Let us now consider the Pontryagin dual group $\widehat{G}$ and dual endomorphism $\widehat{\phi} : \rho \mapsto \rho \circ \phi$ on the $\widehat{G}$.

We shall require the following statement:

LEMMA 5.1. [**13**] *Let* $\phi : G \to G$ *be an endomorphism of an Abelian group* G. *Then the kernel* $\ker \left[\hat{\phi} : \hat{G} \to \hat{G}\right]$ *is canonically isomorphic to the Pontryagin dual of* Coker ϕ.

PROOF. We construct the isomorphism explicitly. Let χ be in the dual of Coker $(\phi : G \to G)$. In that case χ is a homomorphism

$$\chi : G/\operatorname{Im}(\phi) \longrightarrow U(1).$$

There is therefore an induced map

$$\overline{\chi} : G \longrightarrow U(1)$$

which is trivial on Im (ϕ). This means that $\overline{\chi} \circ \phi$ is trivial, or in other words $\hat{\phi}(\overline{\chi})$ is the identity element of $\hat{G}$. We therefore have $\overline{\chi} \in \ker(\hat{\phi})$.

If on the other hand we begin with $\overline{\chi} \in \ker(\hat{\phi})$, then it follows that χ is trivial on Im ϕ, and so $\overline{\chi}$ induces a homomorphism

$$\chi : G/\operatorname{Im}(\phi) \longrightarrow U(1)$$

and χ is then in the dual of Coker ϕ. The correspondence $\chi \leftrightarrow \overline{\chi}$ is clearly a bijection.

□

The following results are also needed

LEMMA 5.2. [**25**] *Let* $L \subset N$ *be* R*-modules and* $g \in R$.
Then
(1)

$$\Big|\frac{N}{gN}\Big| = \Big|\frac{N/L}{g(N/L)}\Big|\Big|\frac{L}{L \cap gN}\Big|$$

(2) If N/L *is finite and the map* $x \to gx$ *is a monomorphism of* N *then*

$$\Big|\frac{N}{gN}\Big| = \Big|\frac{L}{gL}\Big|.$$

Suppose that G as an R- module satisfies the following conditions:

(1) the set of associated primes $Ass(G)$ is finite and consists entirely of non-zero principal ideals of R,

(2) the map $g \to (t^j - 1)g$ is a monomorphism of G for all $j \in \mathbb{N}$ (equivalently, $t^j - 1 \notin \mathfrak{p}$ for all $\mathfrak{p} \in Ass(G)$ and all $j \in \mathbb{N}$),

(3) for each $\mathfrak{p} \in Ass(G)$, $m(\mathfrak{p}) = \dim_{\mathbb{K}(\mathfrak{p})} G_{\mathfrak{p}} < \infty$, where $\mathbb{K}(\mathfrak{p})$ denotes the field of fractions of $R/\mathfrak{p}$ and $G_{\mathfrak{p}} = G \otimes_R \mathbb{K}(\mathfrak{p})$ is the localization of the module G at $\mathfrak{p}$.

LEMMA 5.3. [**25**] *Let* N *be an* R*-module for which* $Ass(N)$ *consists of finitely many non-trivial principal ideals and suppose* $m(\mathfrak{p}) = \dim_{\mathbb{K}(\mathfrak{p})} N_{\mathfrak{p}} < \infty$. *If* $g \in R$ *is such that the map* $x \to gx$ *is a monomorphism of* N, *then* N/gN *is finite.*

If the Pontryagin dual endomorphism $\widehat{\phi}$ is an ergodic finite entropy epimorphism of the compact connected Abelian group $\widehat{G}$ of finite dimension $d \geq 1$ then the endomorphism $\phi : G \to G$ satisfies the conditions (1) - (3) above. Such dual groups $\widehat{G}$ are called solenoids(see [**25, 28**]).

For the dual endomorphism $\widehat{\phi} : \widehat{G} \to \widehat{G}$, we use the following closed periodic point counting formula taken from [**25**, Th. 1.1] and [**1**, Pr. 14]. Let $F_{\widehat{\phi}}(j) = |Fix(\widehat{\phi}^j)|$ denotes the number of points fixed by the endomorphism $\widehat{\phi}^j$. Some obvious conditions such as ergodicity and finite entropy are necessary to ensure that $F_{\widehat{\phi}}(j)$ is finite for all $j \in \mathbb{N}$. Let $\mathcal{P}(\mathbb{K})$ denote the set of finite places of the algebraic number field $\mathbb{K}$. The places of the field $\mathbb{K}$ are the equivalence classes of absolute values on $\mathbb{K}$. When $char(\mathbb{K}) = 0$, the infinite places are the archimedean ones. All other places are said to be finite. Given a finite place of $\mathbb{K}$, there corresponds a unique discrete valuation v whose precise value group is $\mathbb{Z}$. The corresponding normalised absolute value $|\cdot|_v = |\mathcal{R}_v|^{-v(\cdot)}$, where $\mathcal{R}_v$ is the residue class field of v. For any set of places S, we write $|x|_S = \prod_{v\in S} |x|_v$.

Proposition 5.4. [**25**, Th. 1.1], [**1**, Pr. 14] *If $\widehat{\phi} : \widehat{G} \to \widehat{G}$ is an ergodic finite entropy automorphism of a finite dimensional compact connected Abelian group $\widehat{G}$, then there exist algebraic number fields $\mathbb{K}_1, \dots, \mathbb{K}_n$, sets of finite places $P_i \subset \mathcal{P}(\mathbb{K}_i)$ and elements $\xi_i \in \mathbb{K}_i$, no one of which is a root of unity for $i = 1, \dots, n$, such that for any $j \in \mathbb{N}$.*

$$F_{\widehat{\phi}}(j) = \prod_{i=1}^{n} \prod_{v\in P_i} |\xi_i^j - 1|_v^{-1} = \prod_{i=1}^{n} |\xi_i^j - 1|_{P_i}^{-1}. \tag{7}$$

Proof. We outline the major steps in the proof.

Under assumptions of the proposition the number of the periodic points $F_{\widehat{\phi}}(j)$ is finite for all $j \in \mathbb{N}$. Considering abelian group G as $Z[t]$-module and using a straightforward duality argument in Lemma 5.1(or in [**23**, Lemma 7.2]) we have

$$\begin{aligned} F_{\widehat{\phi}}(j) &= |Fix(\widehat{\phi}^j)| = |\operatorname{Ker}(\widehat{\phi}^j - \mathrm{Id}_{\widehat{G}})| = |\widehat{\operatorname{Coker}\,(\phi - \mathrm{Id}_G)}| \\ &= |\operatorname{Coker}\,(\phi^j - \mathrm{Id}_G)| = |G/(\phi^j - 1)G| = |G/(t^j - 1)G|. \end{aligned}$$

The multiplicative set $U = \bigcap_{\mathfrak{p}\in Ass(G)} R - \mathfrak{p}$ has $U \cap ann(a) = \emptyset$ for all non-zero $a \in G$, so the natural map $G \to U^{-1}G$ is a monomorphism. Identifying localizations of R with subrings of $\mathbb{Q}(t)$, the domain $\mathfrak{R} = U^{-1}R = \bigcap_{\mathfrak{p}\in Ass(G)} R_{\mathfrak{p}}$ is a finite intersection of discrete valuation rings and is therefore a principal ideal domain [**24**]. The assumptions of finite entropy and finite topological dimension force $U^{-1}G$ to be a Noeterian $\mathfrak{R}$ - module.Hence, there is a prime filtration

$$\{0\} = G_0 \subset G_1 \subset \cdots \subset G_n = U^{-1}G$$

in which $G_i/G_{i-1} \cong \mathfrak{R}/\mathfrak{q}_i$ for non-trivial primes $\mathfrak{q}_i \subset \mathfrak{R}, 1 \le i \le n$. Moreover, $\mathfrak{p}_i = \mathfrak{q}_i \cap R \in Ass(G)$ for all $1 \le i \le n$. Identifying G with its image in $U^{-1}G$ and intersecting the chain above with G gives a chain

$$\{0\} = L_0 \subset L_1 \subset \cdots \subset L_n = G.$$

Considering this chain of R-modules, for each $1 \le i \le n$ there is an induced inclusion

$$\frac{L_i}{L_{i-1}} \hookrightarrow \frac{G_i}{G_{i-1}} \cong \frac{\mathfrak{R}}{\mathfrak{q}_i} \cong \mathbb{K}(\mathfrak{p}_i) = K_i$$

and $N_i = L_i/L_{i-1}$ may be considered as a fractional ideal of $E_i = R/\mathfrak{p}_i$. Using Lemma 5.2(1),

$$\Big|\frac{L_i}{(t^j-1)L_i}\Big| = \Big|\frac{N_i}{(t^j-1)N_i}\Big|\Big|\frac{L_{i-1}}{L_{i-1}\cap(t^j-1)L_i}\Big|,$$

where $1 \le i \le n$. Let $y \in L_i$, let η denote the image of y in N_i and let ξ_i denote the image of t in E_i. If $(t^j-1)y \in L_{i-1}$ then $(\xi_i^j-1)\eta = 0$. The ergodicity assumption implies $t^j - 1 \notin \mathfrak{p}_i$ so $(\xi_i^j - 1) \neq 0$. Therefore, $\eta = 0$ and $y \in L_{i-1}$. It follows that $L_{i-1} \cap (t^j-1)L_i = (t^j-1)L_{i-1}$ and hence,

$$\Big|\frac{L_i}{(t^j-1)L_i}\Big| = \Big|\frac{N_i}{(t^j-1)N_i}\Big|\Big|\frac{L_{i-1}}{(t^j-1)L_{i-1}}\Big|,$$

Successively applying this formula to each of the modules $L_i, 1 \le i \le n$, gives,

$$|G/(t^j-1)G| = \prod_{i=1}^{n} |N_i/(t^j-1)N_i|$$

Consider now an individual term $|N_i/(t^j-1)N_i|$. Since E_i is a finitely generated domain, [**25**, Th. 1.1][**8**, Th. 4.14] shows that the integral closure D_i of E_i in K_i is a finitely generated Dedekind domain. Therefore, D_i is finitely generated as an E_i-module. We may consider $I_i = D_i \otimes_{E_i} N_i$ as a fractional ideal of D_i. Lemma 5.3 and Lemma 5.2(2) imply that $|N_i/(\xi_i^j-1)N_i| = |I_i/(\xi_i^j-1)I_i|$ (see [**25**]). By considering $I_i/(\xi_i^j-1)I_i$ as a D_i-module, finding a composition series for this module and successively localizing at each of its associated primes to obtain multiplicities, it follows that

$$|I_i/(\xi_i^j-1)I_i| = \prod_{\mathfrak{m}\in Ass(I_i/(\xi_i^j-1)I_i)} q_{\mathfrak{m}}^{\delta_{\mathfrak{m}}(\xi_i,I_i)},$$

where $q_{\mathfrak{m}} = |D_i/\mathfrak{m}|$ and $\delta_{\mathfrak{m}}(\xi_i, I_i) = \dim_{D_i/\mathfrak{m}}(I_i/(\xi_i^j-1)I_i)_{\mathfrak{m}}$. Let

$$P_i = \{\mathfrak{m} \in Spec(D_i) : I_{\mathfrak{m}} \neq K_i\}.$$

It follows that the product above may be taken over all $\mathfrak{m} \in P_i$ to yield the same result. Each localization $(D_i)_{\mathfrak{m}}$ is a distinct valuation ring of K_i and P_i may be identified with a set of finite places of the global field K_i. Hence, since $\delta_{\mathfrak{m}}(\xi_i, D_i) = v_{\mathfrak{m}}(\xi_i^j - 1)$, finally we have

$$|I_i/(\xi_i^j-1)I_i| = \prod_{\mathfrak{m}\in P_i} q_{\mathfrak{m}}^{\delta_{\mathfrak{m}}(\xi_i,D_i)} = \prod_{\mathfrak{m}\in P_i} q_{\mathfrak{m}}^{v_{\mathfrak{m}}(\xi_i^j-1)} = \prod_{\mathfrak{m}\in P_i} |\xi_i^j - 1|_{\mathfrak{m}}^{-1},$$

where $|\cdot|_{\mathfrak{m}}$ is the normalised absolute value arising from $D_{\mathfrak{m}}$. This concludes the proof. □

REMARK 5.5. It is useful to note that $\mathbb{K}_i = \mathbb{Q}(\xi_i)$, $i = 1, \dots, n$. Applying the Artin product formula [**32**] to (7) gives

$$F_{\widehat{\phi}}(j) = \prod_{i=1}^{n} |\xi_i^j - 1|_{P_i^\infty \cup S_i}, \tag{8}$$

where P_i^∞ denotes the set of infinite places of $\mathbb{K}_i$ and $S_i = \mathcal{P}(\mathbb{K}_i) \setminus P_i$. It is also worth noting that [**25**, Rmk. 1] implies that $|\xi_i|_v = 1$ for all $v \in P_i$, $i = 1, \dots, n$, as ϕ is an automorphism.

The following results are needed to have more ready access to the theory of linear recurrence sequences. Relevant background on the connection between linear recurrence sequences and the rationality may be found in the monograph of Everest, van der Poorten, Shparlinski and Ward [**10**].

LEMMA 5.6. *(cf. [**1**]) Let $R(z) = \sum_{n=1}^{\infty} R(\phi^n) z^n$. If $R_\phi(z)$ is rational then $R(z)$ is rational. If $R_\phi(z)$ has analytic continuation beyond its circle of convergence, then so too does $R(z)$. In particular, the existence of a natural boundary at the circle of convergence for $R(z)$ implies the existence of a natural boundary for $R_\phi(z)$.*

PROOF. This follows from the fact that $R(z) = z \cdot R'_\phi(z)/R_\phi(z)$. □

One of the important links between the arithmetic properties of the coefficients of a complex power series and its analytic behaviour is given by the Pólya–Carlson theorem [**3**], [**27**], [**29**].

Pólya–Carlson Theorem. *A power series with integer coefficients and radius of convergence* 1 *is either rational or has the unit circle as a natural boundary.*

For the proof of the main theorem of this section we use the following key result of Bell, Miles and Ward .

LEMMA 5.7 (Lemma 17 in [**1**]). *Let S be a finite list of places of algebraic number fields and, for each $v \in S$, let ξ_v be a non-unit root in the appropriate number field such that $|\xi_v|_v = 1$. Then the function*

$$F(z) = \sum_{n=1}^{\infty} f(n) z^n,$$

where $f(n) = \prod_{v \in S} |\xi_v^n - 1|_v$ for $n \geqslant 1$, has the unit circle as a natural boundary.

The main results of this section are the following counting formulas for the Reidemeister numbers and a Pólya–Carlson dichotomy between rationality and a natural boundary for the analytic behaviour of the Reidemeister zeta function. We follow the method of the proof of Bell, Miles and Ward in [**1**, Theorem 15] for the Artin–Masur zeta function of compact abelian groups automorphisms.

THEOREM 5.8. *Let $\phi : G \to G$ be an automorphism of a countable Abelian group G that is a subgroup of $\mathbb{Q}^d$, where $d \geqslant 1$. Suppose that the group G as $R = \mathbb{Z}[t]$- module satisfies the following conditions:*

(1) the set of associated primes $Ass(G)$ is finite and consists entirely of non-zero principal ideals of the polynomial ring $R = \mathbb{Z}[t]$,

(2) the map $g \to (t^j - 1)g$ is a monomorphism of G for all $j \in \mathbb{N}$ (equivalently, $t^j - 1 \notin \mathfrak{p}$ for all $\mathfrak{p} \in Ass(G)$ and all $j \in \mathbb{N}$),

(3) for each $\mathfrak{p} \in Ass(G)$, $m(\mathfrak{p}) = \dim_{\mathbb{K}(\mathfrak{p})} G_{\mathfrak{p}} < \infty$.

Then there exist algebraic number fields $\mathbb{K}_1, \ldots, \mathbb{K}_n$, sets of finite places $P_i \subset \mathcal{P}(\mathbb{K}_i)$, $S_i = \mathcal{P}(\mathbb{K}_i) \setminus P_i$, and elements $\xi_i \in \mathbb{K}_i$, no one of which is a root of unity for $i = 1, \ldots, n$, such that

$$R(\phi^j) = \prod_{i=1}^{n} \prod_{v \in P_i} |\xi_i^j - 1|_v^{-1} = \prod_{i=1}^{n} |\xi_i^j - 1|_{P_i}^{-1} = \prod_{i=1}^{n} |\xi_i^j - 1|_{P_i^\infty \cup S_i} \tag{9}$$

for all $j \in \mathbb{N}$.

Suppose that the product in (9) *only involves finitely many places and that $|\xi_i|_v \neq 1$ for all v in the set of infinite places P_i^∞ of $\mathbb{K}_i$ and all $i = 1, \ldots, n$.*

Then the Reidemeister zeta function $R_\phi(z)$ is either rational function or has a natural boundary at its circle of convergence, and the latter occurs if and only if $|\xi_i|_v = 1$ for some $v \in S_i$, $1 \leqslant i \leqslant n$.

PROOF. The Reidemeister number of an endomorphism ϕ of an Abelian group G coincides with the cardinality of the quotient group Coker $(\phi - \mathrm{Id}_G) = G/\mathrm{Im}(\phi - \mathrm{Id}_G)$ (or Coker $(\mathrm{Id}_G - \phi) = G/\mathrm{Im}(\mathrm{Id}_G - \phi)$). A straightforward duality argument using Lemma 5.1 shows that

$$R(\phi) = |\,\mathrm{Coker}\ (\phi - \mathrm{Id}_G)| = |\widehat{\mathrm{Coker}\ (\phi - \mathrm{Id}_G)}| = |\,\mathrm{Ker}(\widehat{\phi} - \mathrm{Id}_{\widehat{G}})| = |Fix(\widehat{\phi})|. \tag{10}$$

If the endomorphism $\phi : G \to G$ satisfies the conditions (1) - (3), then the Pontryagin dual endomorphism $\widehat{\phi}$ is an ergodic finite entropy epimorphism of the compact connected Abelian group $\widehat{G}$ of the finite dimension $d \geq 1$ i.e. the Pontryagin dual group $\widehat{G}$ is a solenoid(see [**25, 28**]). Hence the Reidemeister numbers $R(\phi^j)$ and the number of periodic points of the dual map $F_{\widehat{\phi}}(j)$ are finite for all $j \in \mathbb{N}$. By (7), (8) and (10) we have

$$R(\phi^j) = F_{\widehat{\phi}}(j) = \prod_{i=1}^{n} \prod_{v \in P_i} |\xi_i^j - 1|_v^{-1} = \prod_{i=1}^{n} |\xi_i^j - 1|_{P_i}^{-1} = \prod_{i=1}^{n} |\xi_i^j - 1|_{P_i^\infty \cup S_i}. \tag{11}$$

Let $S_i^* = \{v \in S_i : |\xi_i|_v \neq 1\}$, $S_i^{**} = \{v \in S_i : |\xi_i|_v > 1\}$ and let

$$f(j) = \prod_{i=1}^{n} |\xi_i^j - 1|_{S_i \setminus S_i^*}, \quad g(j) = \prod_{i=1}^{n} |\xi_i^j - 1|_{P_i^\infty \cup S_i^*}.$$

So, $R(\phi^j) = f(j)g(j)$ by (9). By the ultrametric property

$$g(j) = \prod_{i=1}^{n} |\xi_i|_{S_i^{**}}^j \cdot |\xi_i^j - 1|_{P_i^\infty}.$$

We can expand the product over infinite places using an appropriate symmetric polynomial to obtain an expression of the form

$$g(j) = \sum_{I \in \mathcal{I}} d_I w_I^j, \tag{12}$$

where $\mathcal{I}$ is a finite indexing set, $d_I \in \{-1, 1\}$ and $w_I \in \mathbb{C}$.

Furthermore, by (12),

$$R_\phi(z) = \exp\left(\sum_{I \in \mathcal{I}} d_I \sum_{j=1}^{\infty} \frac{f(j)(w_I z)^j}{j} \right).$$

If $S_i \setminus S_i^* = \varnothing$ for all $i = 1, \ldots, n$, then $f(j) \equiv 1$, and it follows immediately that the Reidemeister zeta function $R_\phi(z)$ is rational function.

Now suppose that $S_i \setminus S_i^* \neq \varnothing$ for some i. As noted in Lemma 5.6, we need only exhibit a natural boundary at the circle of convergence for

$$\sum_{I \in \mathcal{I}} d_I \sum_{j=1}^{\infty} f(j)(w_I z)^j$$

to exhibit one for $R_\phi(z)$. Moreover, $\limsup_{j\to\infty} f(j)^{1/j} = 1$, so for each $I \in \mathcal{I}$, the series

$$\sum_{j=1}^{\infty} f(j)(w_I z)^j$$

has radius of convergence $|w_I|^{-1}$.

Since $|\xi_i|_v \neq 1$ for all $v \in P_i^\infty$, $i = 1, \dots, n$, there is a dominant term w_J in the expansion (12), for which

$$|w_J| = \prod_{i=1}^{n} |\xi_i|_{S_i^{**}} \prod_{v\in P_i^\infty} \max\{|\xi_i|_v, 1\} = \prod_{i=1}^{n} \prod_{v\in P_i^\infty \cup \mathcal{P}(\mathbb{K}_j)} \max\{|\xi_i|_v, 1\},$$

and $|w_J| > |w_I|$ for all $I \neq J$ (note that $\log|w_J|$ is the topological entropy, as given by [**22**]).

Since $|w_J|^{-1} < |w_I|^{-1}$ for all $I \neq J$, this means that it suffices to show that the circle of convergence $|z| = |w_J|^{-1}$ is a natural boundary for $\sum_{j=1}^{\infty} f(j)(w_I z)^j$. But this is the case precisely when $\sum_{j=1}^{\infty} f(j)z^j$ has the unit circle as a natural boundary, and this has already been dealt with by Lemma 5.7. □

5.1. Examples. To give an example of irrational Reidemeister zeta function let us consider an endomorphism $\phi : g \to 2g$ on the module $\mathbb{Z}[\frac{1}{3}]$ which is an infinitely generated abelian group. We follow the method and the calculations of Everest, Stangoe and Ward in Lemma 4.1 in [**9**] for the Artin–Masur zeta function of the dual compact abelian group endomorphism $\widehat{\phi}$.

LEMMA 5.9. *(cf. Lemma 4.1 of* [**9**] *) The Reidemeister zeta function $R_\phi(z)$ has natural boundary $|z| = \frac{1}{2}$.*

PROOF. The dual compact abelian group $\widehat{\mathbb{Z}[\frac{1}{3}]}$ is a one dimensional solenoid. By (7), (8) and (11) the Reidemeister numbers of iterations of ϕ and the number of periodic points of the dual map $\widehat{\phi}$ are $R(\phi^j) = |Fix(\widehat{\phi}^j)| = F_{\widehat{\phi}}(j) = |2^j - 1| \cdot |2^j - 1|_3$.

Let $\xi(z) = \sum_{n=1}^{\infty} \frac{z^n}{n} |2^n - 1| \cdot |2^n - 1|_3$ so the Reidemeister zeta function $R_\phi(z) = \exp(\xi(z))$. Now

$$\begin{aligned} \xi(z) &= \sum_{n=0}^{\infty} \frac{z^{2n+1}}{2n+1}(2^{2n+1} - 1) + \sum_{n=1}^{\infty} \frac{z^{2n}}{2n}(2^{2n} - 1)|2^{2n} - 1|_3 \\ &= \log\left(\frac{1-z}{1-2z}\right) - \frac{1}{2}\log\left(\frac{1-z^2}{1-4z^2}\right) + \sum_{n=1}^{\infty} \frac{z^{2n}}{2n}(2^{2n} - 1)|2^{2n} - 1|_3. \end{aligned}$$

Notice that

$$|2^n - 1|_3 = |(3-1)^n - 1|_3 = |3^n - n3^{n-1} + \cdots + (-1)^{n-1}3n + (-1)^n - 1|_3 = \\ = \begin{cases} \frac{1}{3}|n|_3 & \text{if } n \text{ is even,} \\ 1 & \text{if } n \text{ is odd.} \end{cases}$$

In particular,

$$|4^n - 1|_3 = |2^{2n} - 1|_3 = \tfrac{1}{3}|2n|_3 = \tfrac{1}{3}|n|_3. \tag{13}$$

Write $\frac{1}{6}\xi_1(z)$ for the last term in an expression for $\xi(z)$ above, so by (13)

$$\xi_1(z) = 3\sum_{n=1}^{\infty}\frac{z^{2n}}{n}(4^n-1)|4^n-1|_3 = \sum_{n=1}^{\infty}\frac{z^{2n}}{n}(4^n-1)|n|_3.$$

Following Lemma 4.1 in [**9**] we shall show that $\xi_1(z)$ has infinitely many logarithmic singularities on the circle $|z| = \frac{1}{2}$, each of which corresponds to a zero of the Reidemeister zeta function $R_\phi(z)$.

Write $3^a\|n$ to mean that $3^a|n$ but $3^{a+1}\nmid n$. Notice that $3^a\|n$ if and only if $|n|_3 = 3^{-a}$. Then ξ_1 may be split up according to the size of $|n|_3$ as

$$\xi_1(z) = \sum_{j=0}^{\infty}\frac{1}{3^j}\sum_{3^j\|n}\frac{z^{2n}}{n}(4^n-1) = \sum_{j=0}^{\infty}\frac{1}{3^j}\eta_j^{(4)}(z),$$

where $\eta_j^{(a)}(z) = \sum_{3^j\|n}\frac{z^{2n}}{n}(a^n-1)$. Then

$$\begin{aligned}\eta_0^{(a)}(z) &= \sum_{3^0\|n}\frac{z^{2n}}{n}(a^n-1) = \sum_{n=1}^{\infty}\frac{z^{2n}}{n}(a^n-1) - \sum_{n=1}^{\infty}\frac{z^{6n}}{3n}(a^{3n}-1)\\ &= \log\left(\frac{1-z^2}{1-az^2}\right) - \frac{1}{3}\log\left(\frac{1-z^6}{1-a^3z^6}\right),\end{aligned}$$

$$\eta_1^{(4)}(z) = \sum_{3^1\|n}\frac{z^{2n}}{n}(4^n-1) = \sum_{3^0\|n}\frac{z^{6n}}{3n}(4^{3n}-1) = \tfrac{1}{3}\eta_0^{(4^3)}(z^3),$$

$$\eta_2^{(4)}(z) = \tfrac{1}{9}\eta_0^{(4^9)}(z^9),$$

and so on. Thus

$$\xi_1(z) = \log\left(\frac{1-z^2}{1-(2z)^2}\right) + 2\sum_{j=1}^{\infty}\frac{1}{9^j}\log\left(\frac{1-(2z)^{2\times 3^j}}{1-z^{2\times 3^j}}\right),$$

so for the Reidemeister zeta function we have

$$|R_\phi(z)| = \left|\frac{1-z}{1-2z}\right| \cdot \left|\frac{1-(2z)^2}{1-z^2}\right|^{1/2} \cdot \left|\frac{1-z^2}{1-(2z)^2}\right|^{1/6} \cdot \prod_{j=1}^{\infty}\left|\frac{1-(2z)^{2\times 3^j}}{1-z^{2\times 3^j}}\right|^{1/3\times 9^j}.$$

It follows that the series defining the Reidemeister zeta function $R_\phi(z)$ has a zero at all points of the form $\frac{1}{2}e^{2\pi ij/3^r}$, $r \geqslant 1$ so $|z| = \frac{1}{2}$ is a natural boundary for the Reidemeister zeta function $R_\phi(z)$. $\square$

References

[1] Jason Bell, Richard Miles, and Thomas Ward, *Towards a Pólya-Carlson dichotomy for algebraic dynamics*, Indag. Math. (N.S.) **25** (2014), no. 4, 652–668, DOI 10.1016/j.indag.2014.04.005. MR3217030

[2] Bachir Bekka, Pierre de la Harpe, and Alain Valette, *Kazhdan's property (T)*, New Mathematical Monographs, vol. 11, Cambridge University Press, Cambridge, 2008. MR2415834

[3] Fritz Carlson, *Über ganzwertige Funktionen* (German), Math. Z. **11** (1921), no. 1-2, 1–23, DOI 10.1007/BF01203188. MR1544479

[4] V. Chothi, G. Everest, and T. Ward, *S-integer dynamical systems: periodic points*, J. Reine Angew. Math. **489** (1997), 99–132. MR1461206

[5] Karel Dekimpe and Gert-Jan Dugardein, *Nielsen zeta functions for maps on infra-nilmanifolds are rational*, J. Fixed Point Theory Appl. **17** (2015), no. 2, 355–370, DOI 10.1007/s11784-014-0165-4. MR3397121

[6] Karel Dekimpe, Sam Tertooy, and Iris Van den Bussche, *Reidemeister zeta functions of low-dimensional almost-crystallographic groups are rational*, Comm. Algebra **46** (2018), no. 9, 4090–4103, DOI 10.1080/00927872.2018.1435792. MR3820618

[7] Pierre Deligne, *La conjecture de Weil. I* (French), Inst. Hautes Études Sci. Publ. Math. **43** (1974), 273–307. MR0340258

[8] David Eisenbud, *Commutative algebra*, Graduate Texts in Mathematics, vol. 150, Springer-Verlag, New York, 1995. With a view toward algebraic geometry. MR1322960

[9] G. Everest, V. Stangoe, and T. Ward, *Orbit counting with an isometric direction*, Algebraic and topological dynamics, Contemp. Math., vol. 385, Amer. Math. Soc., Providence, RI, 2005, pp. 293–302, DOI 10.1090/conm/385/07202. MR2180241

[10] Graham Everest, Alf van der Poorten, Igor Shparlinski, and Thomas Ward, *Recurrence sequences*, Mathematical Surveys and Monographs, vol. 104, American Mathematical Society, Providence, RI, 2003. MR1990179

[11] A. L. Fel′shtyn, *The Reidemeister zeta function and the computation of the Nielsen zeta function*, Colloq. Math. **62** (1991), no. 1, 153–166, DOI 10.4064/cm-62-1-153-166. MR1114629

[12] A. Fel'shtyn, Dynamical zeta functions, Nielsen theory and Reidemeister torsion, Mem. Amer. Math. Soc., **699**, Amer. Math. Soc., Providence, R.I., 2000.

[13] Alexander Fel′shtyn and Richard Hill, *The Reidemeister zeta function with applications to Nielsen theory and a connection with Reidemeister torsion*, K-Theory **8** (1994), no. 4, 367–393, DOI 10.1007/BF00961408. MR1300546

[14] Alexander Fel′shtyn, Richard Hill, and Peter Wong, *Reidemeister numbers of equivariant maps*, Topology Appl. **67** (1995), no. 2, 119–131, DOI 10.1016/0166-8641(94)00011-8. MR1362078

[15] Alexander Fel'shtyn and Jong Bum Lee, *The Nielsen and Reidemeister numbers of maps on infra-solvmanifolds of type (R)*, Topology Appl. **181** (2015), 62–103, DOI 10.1016/j.topol.2014.12.003. MR3303933

[16] Alexander Fel′shtyn and Evgenij Troitsky, *Twisted Burnside-Frobenius theory for discrete groups*, J. Reine Angew. Math. **613** (2007), 193–210, DOI 10.1515/CRELLE.2007.097. MR2377135

[17] A. L. Fel'shtyn and E. V. Troitsky, *Twisted Burnside-Frobenius theory for endomorphisms of polycyclic groups*, Russ. J. Math. Phys. **25** (2018), no. 1, 17–26, DOI 10.1134/S1061920818010028. MR3773874

[18] A. Fel'shtyn, E. Troitsky, M. Zietek, New zeta functions of Reidemeister type and twisted Burnside-Frobenius theory, http://arxiv.org/abs/1804.02874.

[19] David Fried, *The zeta functions of Ruelle and Selberg. I*, Ann. Sci. École Norm. Sup. (4) **19** (1986), no. 4, 491–517. MR875085

[20] David Fried, *Lefschetz formulas for flows*, The Lefschetz centennial conference, Part III (Mexico City, 1984), Contemp. Math., vol. 58, Amer. Math. Soc., Providence, RI, 1987, pp. 19–69. MR893856

[21] Li Li, *On the rationality of the Nielsen zeta function* (English, with English and Chinese summaries), Adv. in Math. (China) **23** (1994), no. 3, 251–256. MR1292755

[22] D. A. Lind and T. Ward, *Automorphisms of solenoids and p-adic entropy*, Ergodic Theory Dynam. Systems **8** (1988), no. 3, 411–419, DOI 10.1017/S0143385700004545. MR961739

[23] Douglas Lind, Klaus Schmidt, and Tom Ward, *Mahler measure and entropy for commuting automorphisms of compact groups*, Invent. Math. **101** (1990), no. 3, 593–629, DOI 10.1007/BF01231517. MR1062797

[24] Hideyuki Matsumura, *Commutative ring theory*, 2nd ed., Cambridge Studies in Advanced Mathematics, vol. 8, Cambridge University Press, Cambridge, 1989. Translated from the Japanese by M. Reid. MR1011461

[25] Richard Miles, *Periodic points of endomorphisms on solenoids and related groups*, Bull. Lond. Math. Soc. **40** (2008), no. 4, 696–704, DOI 10.1112/blms/bdn052. MR2441142

[26] Denis Osin, *Small cancellations over relatively hyperbolic groups and embedding theorems*, Ann. of Math. (2) **172** (2010), no. 1, 1–39, DOI 10.4007/annals.2010.172.1. MR2680416

[27] G. Pólya, *Über gewisse notwendige Determinantenkriterien für die Fortsetzbarkeit einer Potenzreihe* (German), Math. Ann. **99** (1928), no. 1, 687–706, DOI 10.1007/BF01459120. MR1512473

[28] Klaus Schmidt, *Dynamical systems of algebraic origin*, Progress in Mathematics, vol. 128, Birkhäuser Verlag, Basel, 1995. MR1345152

[29] Sanford L. Segal, *Nine introductions in complex analysis*, Revised edition, North-Holland Mathematics Studies, vol. 208, Elsevier Science B.V., Amsterdam, 2008. MR2376066

[30] Jean-Pierre Serre, *Trees*, Springer Monographs in Mathematics, Springer-Verlag, Berlin, 2003. Translated from the French original by John Stillwell; Corrected 2nd printing of the 1980 English translation. MR1954121

[31] S. Smale, *Differentiable dynamical systems*, Bull. Amer. Math. Soc. **73** (1967), 747–817, DOI 10.1090/S0002-9904-1967-11798-1. MR228014

[32] André Weil, *Basic number theory*, Die Grundlehren der mathematischen Wissenschaften, Band 144, Springer-Verlag New York, Inc., New York, 1967. MR0234930

[33] A. V. Zarelua, *On congruences for the traces of powers of some matrices* (Russian, with Russian summary), Tr. Mat. Inst. Steklova **263** (2008), no. Geometriya, Topologiya i Matematicheskaya Fizika. I, 85–105, DOI 10.1134/S008154380804007X; English transl., Proc. Steklov Inst. Math. **263** (2008), no. 1, 78–98. MR2599373

Instytut Matematyki, Uniwersytet Szczecinski, ul. Wielkopolska 15, 70-451 Szczecin, Poland

Email address: `fels@wmf.univ.szczecin.pl, felshtyn@gmail.com`

Instytut Matematyki, Uniwersytet Szczecinski, ul. Wielkopolska 15, 70-451 Szczecin, Poland

Email address: `malwina.zietek@gmail.com`

Contemporary Mathematics
Volume **744**, 2020
https://doi.org/10.1090/conm/744/14980

Rigorous dimension estimates for Cantor sets arising in Zaremba theory

Oliver Jenkinson and Mark Pollicott

In memoriams: Jean Bourgain and Sergiy Kolyada

Abstract. We address the question of the accuracy of bounds used in the study of Zaremba's conjecture. Specifically, we establish rigorous estimates on the Hausdorff dimension of certain Cantor sets which arise in the analysis of Zaremba's conjecture in the work of Bourgain, Kontorovich and Huang.

1. Introduction

Given any rational number $\frac{p}{q} \in (0,1)$ a simple application of Euclid's algorithm shows there exist coefficients $a_1, \cdots, a_n \in \mathbb{N}$ such that

$$\frac{p}{q} = [a_1, \cdots, a_n] := \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cdots + \frac{1}{a_n}}}}$$

(see e.g. [**15**, Thm. 161]). Given a finite subset $A \subset \mathbb{N}$, however, a natural question is to enquire as to what restriction is imposed on the denominators of such rational numbers in the case where $a_1, \cdots, a_n \in A$, in other words to study the corresponding denominator set

$$Q_A = \left\{ q \in \mathbb{N} : \exists p \in \mathbb{N}, a_1, \cdots, a_n \in A \text{ such that } \frac{p}{q} = [a_1, \cdots, a_n] \right\}.$$

More specifically, Zaremba [**28**] conjectured that when $A = \{1,2,3,4,5\}$, all natural numbers occur as denominators q for suitable choices of $a_1, \cdots, a_n \in A$, i.e. that $Q_{\{1,2,3,4,5\}} = \mathbb{N}$.

The choice of numbers up to 5 in this conjecture is natural, since the corresponding result fails for the smaller set $A = \{1,2,3,4\}$, where for example it is known that the numbers 6, 54 and 150 do not lie in the denominator set $Q_{\{1,2,3,4\}}$ (see [**23**, p. 193]).

2010 *Mathematics Subject Classification.* Primary 11K55, 37C30; Secondary 37F35, 11A55.

Key words and phrases. Hausdorff Dimension, Transfer Operators, Thermodynamic Formalism, Zaremba Conjectures.

The first author was partially supported by EPSRC grant EP/L02246X/1.

The second author was partially supported by EPSRC grant EP/M001903/1 and ERC grant Resonances/833802.

The original motivation of Zaremba to study this problem was related to numerical integration and the use of the method of "good lattice points". Although Zaremba's conjecture remains open, there is various numerical evidence supporting it (see e.g. the discussion in [**23**, §2]); indeed in the article [**4**], the authors cite work of Borosh showing that all the denominators $q \le 10^4$ occur in $Q_{\{1,2,3,4,5\}}$, and quote Knuth as having established the same result in the range $10^4 \le q \le 3.2 \times 10^6$.

In a significant recent paper, Bourgain & Kontorovich [**5**] showed that for the larger set $A = \{1, 2, \cdots, 50\}$, the corresponding denominator set Q_A has *density one* as a subset of $\mathbb{N}$, in other words

$$\lim_{N \to +\infty} \frac{Q_A \cap \{1, \ldots, N\}}{N} = 1.$$

There is known to be a close connection between this kind of problem and the *Hausdorff dimension* of certain related sets. For a finite subset $A \subset \mathbb{N}$, let E_A denote the set of all $x \in (0, 1)$ such that the digits $a_1(x), a_2(x), \ldots$ in the (infinite) continued fraction expansion

$$x = [a_1(x), a_2(x), a_3(x), \ldots] = \cfrac{1}{a_1(x) + \cfrac{1}{a_2(x) + \cfrac{1}{a_3(x) + \cdots}}}$$

all belong to A. Sets of the form E_A are said to be of *bounded type* (see e.g. [**23**, **26**]); in particular they are Cantor sets, and study of their Hausdorff dimension has attracted significant attention (see e.g. [**7–10**, **13**, **14**, **16**, **17**, **20–22**]).

In the context of the Zaremba conjecture, the following result of Huang [**18**, **19**] illustrates the connection with bounds on the Hausdorff dimension of bounded type sets E_A:

Theorem 1 (Huang). *For the set $A = \{1, 2, 3, 4, 5\}$, the corresponding denominator set Q_A has density one in $\mathbb{N}$ provided* $\dim(E_A) > \frac{5}{6}$.

In particular, Huang's theorem represents an improvement on the above result of Bourgain-Kontorovich (in that the set $\{1, 2, \ldots, 50\}$ is replaced by the smaller set $\{1, 2, 3, 4, 5\}$), provided the lower bound $\dim(E_A) > \frac{5}{6}$ does indeed hold. In fact Huang [**18**, **19**] cites as evidence of this bound a paper of the first author [**20**], where the techniques of [**21**] were used to give a non-rigorous indication that $\dim(E_A) \approx 0.8368 > 0.8333\ldots = \frac{5}{6}$. Although the method introduced in [**21**] yielded high quality empirical approximations, it is only in our more recent paper [**22**] that effective techniques have been introduced for converting these heuristics into a rigorous proof of the quality of a specific computation. In view of the conditional nature of Huang's Theorem 1, and the recent availability of techniques potentially capable of rendering rigorous the heuristic estimate of $\dim(E_A)$ in [**20**], in this paper we employ the technology of [**22**] in order to rigorously prove the following:

Theorem 2. *If $A = \{1, 2, 3, 4, 5\}$ then* $\dim(E_A) > \frac{5}{6}$.

Indeed in §5 we give a rigorous proof, stated as Theorem 8, of a significantly more accurate estimate on $\dim(E_{\{1,2,3,4,5\}})$. Combining Theorems 1 and 2 we deduce the following unconditional version of Huang's Theorem.

Corollary 1 (after Huang). *For the set* $A = \{1, 2, 3, 4, 5\}$*, the corresponding denominator set* Q_A *has density one in* $\mathbb{N}$.

A stronger conjecture due to Hensley [**17**, Conj. 3, p. 16] was that provided $\dim(E_A) > \frac{1}{2}$ then every sufficiently large natural number occurs as a corresponding denominator, i.e. that Q_A contains all sufficiently large natural numbers. However, Bourgain & Kontorovich [**5**] indicated that $A = \{2, 4, 6, 8, 10\}$ provides a counterexample to this conjecture, noting that in this case Q_A does not contain any natural numbers which are equal to 3 (mod 4), and that moreover $\dim(E_A) \approx 0.517 > 1/2$ (see [**5**, p. 139]). Their approximation to $\dim(E_{\{2,4,6,8,10\}})$, using an implementation of the algorithm of [**21**], is a heuristic one, in the spirit of the empirical computations in [**20, 21**] rather than the rigorously validated version of [**22**]. In view of the importance of this Bourgain-Kontorovich counterexample to Hensley's conjecture, it is of interest to rigorously establish the lower bound on the dimension of $E_{\{2,4,6,8,10\}}$ (which we present in §4 as Theorem 7):

Theorem 3. *If* $A = \{2, 4, 6, 8, 10\}$ *then* $\dim(E_A) > \frac{1}{2}$.

In particular, this confirms the assertion of [**5**, p. 139], yielding:

Corollary 2 (after Bourgain-Kontorovich). *The set* $A = \{2, 4, 6, 8, 10\}$ *provides a counterexample to the Hensley conjecture.*

Finally, we recall that Bourgain & Kontorovich proved [**5**, Thm. 1.26] the existence of $h \in \mathbb{N}$ such that there are infinitely many prime numbers d which have a primitive root $b \pmod d$ with the property that the partial quotients of the rational b/d are bounded by h, and they indicated that h could be chosen to equal 51. The following sharpening of this result due to Huang [**19**, Cor. 1.1.12] is reliant on a lower bound for the Hausdorff dimension of the Cantor set $E_{\{1,2,3,4,5,6\}}$:

Theorem 4 (Huang). *If* $\dim(E_{\{1,2,3,4,5,6\}}) > \frac{19}{22}$ *then there are infinitely many prime numbers* d *which have a primitive root* $b \pmod d$ *such that the partial quotients of* b/d *are* ≤ 7.

Although Huang indicates that $\dim(E_{\{1,2,3,4,5,6\}}) > \frac{19}{22}$ is true, citing the empirical approximation $\dim(E_{\{1,2,3,4,5,6\}}) \approx 0.8676 > 0.86363636\ldots = \frac{19}{22}$ of [**20**], there was no rigorous proof of this result. Once again, therefore, there is considerable interest in rendering Theorem 4 an unconditional result by providing a rigorous validation of the Hausdorff dimension bound. This we do in our third main result:

Theorem 5. *If* $A = \{1, 2, 3, 4, 5, 6\}$ *then* $\dim(E_A) > \frac{19}{22}$.

In fact we give a rigorous proof of a significantly more accurate estimate on $\dim(E_{\{1,2,3,4,5,6\}})$ as Theorem 9 in §6. A corollary of Theorem 5 is the unconditional analogue of Huang's Theorem 4:

Corollary 3 (after Huang). *There are infinitely many prime numbers* d *which have a primitive root* $b \pmod d$ *such that the partial quotients of* b/d *are* ≤ 7.

The organisation of this article is as follows. After some preliminaries in §2 on Hausdorff dimension and the thermodynamic underpinnings of our computational approach, in §3 we describe (after [**22**]) the way in which these computations can be converted into rigorous effective bounds. In §4 we prove that the Hausdorff dimension of $E_{\{2,4,6,8,10\}}$ is greater than 1/2, in §5 we establish a rigorous bound on $\dim(E_{\{1,2,3,4,5\}})$ which in particular shows this dimension to be larger than 5/6, and in §6 we rigorously approximate the dimension of $E_{\{1,2,3,4,5,6\}}$, which in particular implies it is larger than 19/22.

2. Preliminaries

In this section we collect a number of results (see also [**21**, **22**]) which underpin our algorithm for approximating Hausdorff dimension.

We begin by recalling some results for continued fractions. For a non-empty finite subset $A \subset \mathbb{N}$, let E_A denote the set of all $x \in (0,1)$ such that the digits $a_1(x), a_2(x), \ldots$ in the continued fraction expansion

$$x = [a_1(x), a_2(x), a_3(x), \ldots] = \cfrac{1}{a_1(x) + \cfrac{1}{a_2(x) + \cfrac{1}{a_3(x)+\cdots}}}$$

all lie in A. Equivalently, if

$$T_n(x) := (n+x)^{-1}$$

then E_A is the smallest non-empty closed set satisfying the self-similarity condition

$$E_A = \cup_{n\in A} T_n(E_A)\,.$$

The *Gauss map*

$$T(x) = \frac{1}{x} \pmod 1$$

is such that $T \circ T_n$ is the identity map for each n, and all of the sets E_A satisfy $T(E_A) = E_A$.

Each of the sets $E_A \subset [0,1]$ is a Cantor set of zero Lebesgue measure, and a natural way to describe their size is via Hausdorff dimension.

DEFINITION 1. For a general set $E \subset \mathbb{R}$, if we define

$$H^\delta_\varepsilon(E) := \inf\left\{ \sum_i \operatorname{diam}(U_i)^\delta : \mathcal{U} = \{U_i\} \text{ is an open cover of } E \right.$$
$$\left. \text{such that each } \operatorname{diam}(U_i) \le \varepsilon \right\},$$

and $H^\delta(E) := \lim_{\varepsilon\to 0} H^\delta_\varepsilon(E)$, then the *Hausdorff dimension* of E, denoted $\dim(E)$, is defined to be the infimum of the set $\{\delta : H^\delta(E) = 0\}$.

For the sets E_A, their Hausdorff dimension coincides with their *box dimension* (see e.g. [**12**]).

For a general continuous function $f : E_A \to \mathbb{R}$, its *pressure* $P(f)$ is defined to be

$$P(f) = \lim_{n\to+\infty} \frac{1}{n} \log\left(\sum_{\substack{T^n x = x \\ x \in E_A}} e^{f(x)+f(Tx)+\ldots+f(T^{n-1}x)} \right),$$

and making the particular choice $f = -s\log|T'|$ leads to an important characterisation of the Hausdorff dimension of E_A (see [**3**, **6**, **12**, **24**]):

LEMMA 1. *The function $\mathbb{R} \to \mathbb{R}$ defined by $s \mapsto P(-s\log|T'|)$ is strictly decreasing, and its unique zero is precisely the Hausdorff dimension* $\dim(E_A)$.

For $s \in \mathbb{R}$, and finite $A \subset \mathbb{N}$, define the *transfer operator* $\mathcal{L}_{A,s}$ by

$$\mathcal{L}_{A,s} f(x) = \sum_{n \in A} \frac{f(T_n x)}{(n+x)^{2s}} .$$

This operator is known to leave invariant a number of natural function spaces, notably the Hilbert Hardy spaces considered below, or for example the Banach space of Lipschitz functions on $[0,1]$. On these spaces the value $e^{P(-s\log|T'|)}$ is an eigenvalue of strictly largest modulus, and is a simple eigenvalue. Consequently, Lemma 1 implies that the Hausdorff dimension of E_A is the unique value $s \in \mathbb{R}$ such that $\mathcal{L}_{A,s}$ has spectral radius equal to 1.

When acting on suitable Hilbert Hardy spaces, the *trace* $\mathrm{tr}(\mathcal{L}_{A,s}^n)$ of each n-th power $\mathcal{L}_{A,s}^n$ is given (see [**21**, **25**]) by

$$\mathrm{tr}(\mathcal{L}_{A,s}^n) = \sum_{\underline{i} \in A^n} \frac{|T_{\underline{i}}'(z_{\underline{i}})|^s}{1 - T_{\underline{i}}'(z_{\underline{i}})} = \sum_{\underline{i} \in A^n} \frac{\prod_{j=0}^{n-1} T^j(z_{\underline{i}})^{2s}}{1 - (-1)^n \prod_{j=0}^{n-1} T^j(z_{\underline{i}})^2} , \tag{1}$$

where the point $z_{\underline{i}}$ is the unique fixed point in $(0,1)$ of the n-fold composition $T_{\underline{i}} = T_{i_1} \circ T_{i_2} \circ \cdots \circ T_{i_n}$ (and hence a period-n point for the Gauss map T), and in particular is a quadratic irrational. The function defined on the complex disc $|z| < e^{-P(-s\log|T'|)}$ (i.e. the disc of convergence of $\sum_{n=1}^{\infty} \frac{z^n}{n}\mathrm{tr}(\mathcal{L}_{A,s}^n)$) by

$$\Delta(z,s) = \exp\left(- \sum_{n=1}^{\infty} \frac{z^n}{n} \mathrm{tr}(\mathcal{L}_{A,s}^n) \right) \tag{2}$$

extends by analytic continuation to an entire function of $\mathbb{C}$, called the *determinant* of $\mathcal{L}_{A,s}$.

When acting on suitable Hilbert Hardy spaces, the eigenvalues of $\mathcal{L}_{A,s}$ are precisely the reciprocals of the zeros of its determinant. In particular, the zero of the function $\Delta(s,\cdot)$ with smallest modulus is $e^{-P(-s\log|T'|)}$, therefore the Hausdorff dimension of E_A is precisely the value of s such that 1 is the zero of minimum modulus of $\Delta(s,\cdot)$.

In fact, when $\mathcal{L}_{A,s}$ acts on such a space of holomorphic functions, its approximation numbers decay at an exponential rate (see [**22**, Cor. 2]), so that $\mathcal{L}_{A,s}$ belongs to an exponential class (cf. [**1**, **2**]) and is in particular a trace class operator, from which the existence and above properties of trace and determinant follow (see [**27**]).

This allows us (cf. [**21**, **22**]) to write $\Delta(z,s)$ as the series $\Delta(z,s) = 1 + \sum_{n=1}^{\infty} \delta_n(s) z^n$, and then set $z=1$ to define the *dimension determinant* $\mathfrak{D}$ by

$$\mathfrak{D}(s) := \Delta(1,s) = 1 + \sum_{n=1}^{\infty} \delta_n(s) ,$$

a holomorphic function which is known to be entire (see [**21**, **25**]). Solutions s of

$$0 = 1 + \sum_{n=1}^{\infty} \delta_n(s) = \mathfrak{D}(s) \tag{3}$$

are such that the value 1 is an eigenvalue for the operator $\mathcal{L}_{A,s}$, and in particular the largest real zero of $\mathfrak{D}$ is precisely the dimension $\dim(E_A)$ (cf. Proposition 1), being the value of s such that 1 is the leading eigenvalue (i.e. of maximum modulus) for the operator $\mathcal{L}_{A,s}$.

The coefficients $\delta_n(s)$ are computable (to arbitrary precision, for a given s) in terms of those periodic points of $T|_{E_A}$ whose period is $\leq n$, using the formula (1). Therefore, for any given $N \in \mathbb{N}$, we may define $\mathfrak{D}_N$ by

$$\mathfrak{D}_N(s) := 1 + \sum_{n=1}^{N} \delta_n(s)\,, \tag{4}$$

so that a solution s_N to the equation $\mathfrak{D}_N(s) = 0$ will be an approximate solution to (3), and the smaller $\sum_{n=N+1}^{\infty} \delta_n(s)$ is the better this approximation will be. In what follows, we use rigorous upper bounds on (the absolute value of) $\sum_{n=N+1}^{\infty} \delta_n(s)$ to yield rigorous estimates on $|s_N - \dim(E_A)|$.

3. Bounding dimension determinant coefficients

We now begin the serious task of converting these theoretical estimates into practical bounds that can be used to complete the proofs of the results stated in the introduction. The key point is that we can employ a number of technical innovations introduced in [**22**] in order to make estimates both effective and rigorous.

Let $A \subset \mathbb{N}$ be finite. An open disc $D \subset \mathbb{C}$ is said to be *admissible (for A)* if $\cup_{i\in A} T_i(D) \subset D$.

For an admissible disc D of radius ϱ, centred at c, let D' be the smallest disc, concentric with D, such that $\cup_{i\in A} T_i(D) \subset D'$, and let ϱ' denote the radius of D'. The associated *contraction ratio* $\theta = \theta_{A,D}$ is then defined as

$$\theta = \theta_{A,D} := \frac{\varrho'}{\varrho}\,.$$

Introducing the notation

$$E_n(\theta) := \frac{\theta^{n(n+1)/2}}{\prod_{i=1}^{n}(1-\theta^i)}\,, \tag{5}$$

we note the super-exponential decay $E_n(\theta) = O(\theta^{\frac{n^2}{2}})$ as $n \to \infty$.

Definition 2. The *Hilbert Hardy space* $H^2(D)$ consists of those functions f which are holomorphic on D such that $\|f\|^2 := \sup_{r<\varrho} \int_0^1 |f^*(c + re^{2\pi it})|^2\, dt < \infty$, with inner product given by

$$(f, g) = \int_0^1 f^*(c + \varrho e^{2\pi it})\overline{g^*(c + \varrho e^{2\pi it})}\, dt\,,$$

where f^* and g^* denote the respective non-tangential limit functions of f and g.

The monomials

$$m_k(z) = \varrho^{-k}(z - c)^k \tag{6}$$

constitute an orthonormal basis of $H^2(D)$.

Admissibility of D ensures that for $s \in \mathbb{R}$, the transfer operator $\mathcal{L}_{A,s}$ preserves $H^2(D)$. In particular,

$$\mathcal{L}_{A,s}(m_k)(z) = \sum_{j\in A} \frac{(T_j(z)-c)^k}{\varrho^k (z+j)^{2s}},$$

and we may use numerical integration to explicitly compute (to arbitrary precision) the norm $\|\mathcal{L}_{A,s}(m_k)\|$ as

$$\|\mathcal{L}_{A,s}(m_k)\|^2 = \int_0^1 \left| \sum_{j\in A} \frac{(T_j(\gamma(t))-c)^k}{\varrho^k(\gamma(t)+j)^{2s}} \right|^2 dt, \tag{7}$$

where $\gamma(t) = c + \varrho e^{2\pi i t}$.

For $j \in A$ the functions

$$w_{j,s}(z) = \frac{1}{(z+j)^{2s}}$$

are holomorphic on the admissible disc D, and we use their *uniform* norms

$$\|w_{j,s}\|_\infty = \sup_{z\in D} |w_{j,s}(z)|,$$

together with the contraction ratio θ, to define the constant

$$K_s = K_{s,A,D} := \frac{\sum_{j\in A} \|w_{j,s}\|_\infty}{\theta\sqrt{1-\theta^2}}. \tag{8}$$

For $s \in \mathbb{R}$ and $n, Q, M, N \in \mathbb{N}$ with $n \le Q \le M \le N$, if we introduce the quantities

$$\alpha_{n,N,+}(s) := \left(\sum_{k=n-1}^{N} \|\mathcal{L}_{A,s}(m_k)\|^2 + \left(\sum_{j\in A} \|w_{j,s}\|_\infty \right)^2 \frac{\theta^{2(N+1)}}{1-\theta^2} \right)^{1/2}, \tag{9}$$

$$\beta^{M,-}_{l,N,+}(s) := \sum_{i_1<\ldots<i_l\le M} \prod_{j=1}^{l} \alpha_{i_j,N,+}(s), \tag{10}$$

$$J_{Q,N,s} := K_s \left(1 + \theta^{2(N+2-Q)}\right)^{1/2}, \tag{11}$$

$$\beta^{M,+}_{n,N,+}(s) := \beta^{M,-}_{n,N,+}(s) + \sum_{l=0}^{n-1} J^{n-l}_{Q,N,s}\, \beta^{M,-}_{l,N,+}(s)\, \theta^{M(n-l)} E_{n-l}(\theta). \tag{12}$$

then the following bound was established in [**22**].

Theorem 6. *Let $A \subset \mathbb{N}$ be finite, and D an admissible disc, with contraction ratio $\theta = \theta_{A,D}$. If $s \in \mathbb{R}$, and $Q, M, N \in \mathbb{N}$ with $n \le Q \le M \le N$, then the dimension determinant coefficients $\delta_n(s)$ satisfy*

$$|\delta_n(s)| \le \min\left(K_s^n E_n(\theta)\ ,\ \beta_{n,N,+}^{M,+}(s)\right) .$$

Remark 1. Theorem 6 was proved in [**22**] using the theory of approximation numbers in Hilbert space. The inequality $|\delta_n(s)| \le K_s^n E_n(\theta)$ from Theorem 6 is referred to as the *Euler bound*, acknowledging Euler's work [**11**] on the identity $E_n(\theta) = \sum_{i_1<\ldots<i_n} \theta^{i_1+\cdots+i_n}$. The term $K_s^n E_n(\theta)$ has a simple closed form, and is $O(\gamma^{n^2})$ as $n \to \infty$ for any $\gamma \in (\theta^{1/2}, 1)$, though the constant $K_s = K_{s,A,D}$ may be large enough (if A is large) to render the tail estimate $|\sum_{n>Q} \delta_n(s)| \le \sum_{n>Q} K_s^n E_n(\theta)$ insufficiently sharp if Q is chosen to be small. The terms $\beta_{n,N,+}^{M,+}(s)$, referred to as *upper computed Taylor bounds* in [**22**], have the virtue of being readily computable to arbitrary precision, but are not available in closed form; their utility, therefore, is in bounding $|\delta_n(s)|$ for $n \le Q$, where Q is chosen so that the tail estimate derived from the Euler bound is sufficiently sharp. In practice M and N will be chosen so that $\beta_{n,N,+}^{M,+}(s)$ agrees with $\beta_{n,N,+}^{M,-}(s)$ (which is given by a notably simpler formula) to very high precision (e.g. several hundred decimal places), i.e. the more complicated term $\sum_{l=0}^{n-1} J_{Q,N,s}^{n-l}\, \beta_{l,N,+}^{M,-}(s)\, \theta^{M(n-l)} E_{n-l}(\theta)$ in (12) effectively plays no computational role; similarly, $\alpha_{n,N,+}(s)$ will in practice agree with $(\sum_{k=n-1}^{N} \|\mathcal{L}_{A,s}(m_k)\|^2)^{1/2}$ to very high precision, so that the term $(\sum_{j\in A} \|w_{j,s}\|_\infty)^2 \frac{\theta^{2(N+1)}}{1-\theta^2}$ in (9) effectively plays no computational role.

4. The Hausdorff dimension of $E_{\{2,4,6,8,10\}}$ is greater than $1/2$

Motivated by the work of Bourgain & Kontorovich [**5**] described in §1, specifically [**5**, p. 139] (see also [**23**, Lem. 2.20]), our aim in this section will be to provide a rigorous proof of the fact that the Hausdorff dimension of $E_{\{2,4,6,8,10\}}$ is greater than 1/2, a result which heretofore has enjoyed a folklore status, based on convincing but non-rigorous numerical work.

Our approach is motivated by the following observation:

Proposition 1. *For any finite alphabet A, if $s_0 \in \mathbb{R}$ is such that the corresponding dimension determinant $\mathfrak{D} = \mathfrak{D}_A$ satisfies $\mathfrak{D}(s_0) < 0$, then $dim(E_A) > s_0$.*

Proof. The method is to show firstly that $\mathfrak{D}$ cannot have real zeros that are larger than $\dim(E_A)$, so that $\mathfrak{D}$, being a continuous function, does not change sign on the interval $(\dim(E_A), \infty)$, and secondly that the derivative $\mathfrak{D}'(s)$ is strictly positive at its zero $s = \dim(E_A)$. This then implies that $\mathfrak{D}$ is strictly positive on $(\dim(E_A), \infty)$, or in other words the desired result that if $\mathfrak{D}(s_0) < 0$ then necessarily $\dim(E_A) > s_0$.

To show that if $s>\dim(E_A)$ then $\mathfrak{D}(s)\neq 0$, recall that $s\mapsto p(s)=P(-s\log|T'|)$ is strictly decreasing on $\mathbb{R}$, with $s=\dim(E_A)$ its unique zero. Therefore $s\mapsto z_1(s)=e^{-p(s)}$, which is the zero of minimum modulus of $\Delta(\cdot,s)$, is a strictly increasing function. In particular, $z_1(\dim(E_A))=1$, so if $s>\dim(E_A)$ then $z_1(s)>1$; thus all zeros of $\Delta(\cdot,s)$ must have modulus strictly larger than 1. Therefore in particular the equation $\Delta(1,s)=0$ has no solutions for $s>\dim(E_A)$, i.e. the equation $\mathfrak{D}(s)=0$ has no solutions for $s>\dim(E_A)$, i.e. $\mathfrak{D}$ has no zeros that are strictly larger than $\dim(E_A)$.

To complete the proof it remains to show that the derivative $\mathfrak{D}'(s)$ is strictly positive at $s=\dim(E_A)$. To see this we use the infinite product

$$\Delta(z,s)=\prod_{r=1}^{\infty}(1-z\lambda_r(s)),$$

where $\lambda_r(s)$ are the eigenvalues of $\mathcal{L}_{A,s}$, listed according to algebraic multiplicity, and ordered so that their absolute values are non-increasing, with in particular $\lambda_1(s)>|\lambda_r(s)|$ for all $r\geq 2$ (since the leading eigenvalue $\lambda_1(s)$ is simple).

If $\Gamma(s):=\prod_{r=2}^{\infty}(1-\lambda_r(s))$ then $\mathfrak{D}(s)=(1-\lambda_1(s))\Gamma(s)$, so

$$\mathfrak{D}'(s)=-\lambda_1'(s)\Gamma(s)+(1-\lambda_1(s))\Gamma'(s)\,,$$

and since $\lambda_1(\dim(E_A))=1$ then

$$\mathfrak{D}'(\dim(E_A))=-\lambda_1'(\dim(E_A))\Gamma(\dim(E_A)).$$

But $s\mapsto\lambda_1(s)=e^{p(s)}$ is strictly decreasing, so $-\lambda_1'(\dim(E_A))>0$, and therefore it remains to show that $\Gamma(\dim(E_A))>0$. For this, note that if $s\in\mathbb{R}$ (in particular if $s=\dim(E_A)$), the coefficients in the power series expansion of $\Delta(z,s)$ are all real, by (1). Therefore non-real zeros of Δ arise as conjugate pairs, both with the same multiplicity. Multiplying out those factors in the product representation of Γ corresponding to conjugate pairs, we see that $\Gamma(\dim(E_A))$ is an infinite product of strictly positive terms (since $|\lambda_r(\dim(E_A))|<1$ for each $r\geq 2$). The sequence of terms converges to 1, since $|\lambda_r(\dim(E_A))|\to 0$, therefore the infinite product converges to a strictly positive value. That is, $\Gamma(\dim(E_A))>0$, as required. □

Having established Proposition 1, our strategy for proving that $\dim(E_{\{2,4,6,8,10\}})>1/2$ will be to show that $\mathfrak{D}(1/2)<0$ for the corresponding dimension determinant $\mathfrak{D}=\mathfrak{D}_{\{2,4,6,8,10\}}$.

In view of the central role of the value $s=1/2$ in this section, we shall write

$$\mathcal{L}_A:=\mathcal{L}_{A,1/2}\,,$$

and

$$\delta_n:=\delta_n(1/2)\,,$$

so that

$$\mathfrak{D}(1/2)=1+\sum_{n=1}^{\infty}\delta_n\,.$$

It will turn out to be sufficient to work with Gauss map orbits of periods 1, 2 and 3, and in Lemmas 2, 3, 4 below we record exact formulae for the corresponding traces of the operator $\mathcal{L}_A$.

Lemma 2. *If* $A = \{2,4,6,8,10\}$ *then* $u_1 = \mathrm{tr}(\mathcal{L}_A)$ *is given by the exact formula*

$$\begin{aligned} u_1 &= \frac{\sqrt{2}-1}{1+\left(\sqrt{2}-1\right)^2} + \frac{\sqrt{5}-2}{1+\left(\sqrt{5}-2\right)^2} + \frac{\sqrt{10}-3}{1+\left(\sqrt{10}-3\right)^2} \\ &\quad + \frac{\sqrt{17}-4}{1+\left(\sqrt{17}-4\right)^2} + \frac{\sqrt{26}-5}{1+\left(\sqrt{26}-5\right)^2} \\ &= \frac{1}{2\sqrt{2}} + \frac{1}{2\sqrt{5}} + \frac{1}{2\sqrt{10}} + \frac{1}{2\sqrt{17}} + \frac{1}{2\sqrt{26}}\,. \end{aligned}$$

Proof. From (1), $u_1 = \sum_{n\in A} z_n/(1+z_n^2)$, where

$$z_n = [n,n,n,\ldots] = \sqrt{k_n^2+1} - k_n\,,$$

for $k_n = n/2$, and the result follows. □

Lemma 3. *If* $A = \{2,4,6,8,10\}$ *then* $u_2 = \frac{1}{2}\mathrm{tr}(\mathcal{L}_A^2)$ *is given by the exact formula*

$$\begin{aligned} u_2 = \frac{1}{2}\Bigg(&\frac{3-2\sqrt{2}}{1-\left(3-2\sqrt{2}\right)^2} + \frac{19-6\sqrt{10}}{1-\left(19-6\sqrt{10}\right)^2} + \frac{33-8\sqrt{17}}{1-\left(33-8\sqrt{17}\right)^2} + \frac{51-10\sqrt{26}}{1-\left(51-10\sqrt{26}\right)^2} \\ &+ \frac{3\left(9-4\sqrt{5}\right)}{1-\left(9-4\sqrt{5}\right)^2} + \frac{2\left(17-12\sqrt{2}\right)}{1-\left(17-12\sqrt{2}\right)^2} + \frac{2\left(7-4\sqrt{3}\right)}{1-\left(7-4\sqrt{3}\right)^2} + \frac{2\left(5-2\sqrt{6}\right)}{1-\left(5-2\sqrt{6}\right)^2} \\ &+ \frac{19-6\sqrt{10}}{1-\left(19-6\sqrt{10}\right)^2} + \frac{2\left(31-8\sqrt{15}\right)}{1-\left(31-8\sqrt{15}\right)^2} + \frac{2\left(11-2\sqrt{30}\right)}{1-\left(11-2\sqrt{30}\right)^2} + \frac{2\left(25-4\sqrt{39}\right)}{1-\left(25-4\sqrt{39}\right)^2} \\ &+ \frac{2\left(13-2\sqrt{42}\right)}{1-\left(13-2\sqrt{42}\right)^2} + \frac{2\left(41-4\sqrt{105}\right)}{1-\left(41-4\sqrt{105}\right)^2} + \frac{2\left(21-2\sqrt{110}\right)}{1-\left(21-2\sqrt{110}\right)^2}\Bigg)\,. \end{aligned}$$

Proof. From (1),

$$u_2 = \frac{1}{2}\sum_{(m,n)\in A^2} \frac{z_{m,n}}{1-z_{m,n}^2}\,,$$

where it can be shown that

$$z_{m,n} = k_{m,n} - \sqrt{k_{m,n}^2 - 1}\,,$$

for

$$k_{m,n} = 1 + \frac{mn}{2}\,.$$

Note that $z_{4,4} = z_{2,8} = z_{8,2} = 9-4\sqrt{5}$, contributing the term with coefficient 3 in the above expression for u_2. Otherwise the four remaining fixed points contribute the terms with coefficient 1, and the 9 remaining period-2 orbits contribute the terms with coefficient 2 (since $z_{m,n} = z_{n,m}$). □

LEMMA 4. *If* $A = \{2,4,6,8,10\}$ *then* $u_3 = \frac{1}{3}\mathrm{tr}(\mathcal{L}_A^3)$ *is given by the exact formula*

$$\begin{aligned}
u_3 \;=\; & \frac{1}{3}\Bigg(\frac{5\sqrt{2}-7}{1+\left(5\sqrt{2}-7\right)^2}+\frac{17\sqrt{5}-38}{1+\left(17\sqrt{5}-38\right)^2}+\frac{37\sqrt{10}-117}{1+\left(37\sqrt{10}-117\right)^2}\\
&+\frac{65\sqrt{17}-268}{1+\left(65\sqrt{17}-268\right)^2}+\frac{101\sqrt{26}-515}{1+\left(101\sqrt{26}-515\right)^2}+\frac{3\left(\sqrt{145}-12\right)}{1+\left(\sqrt{145}-12\right)^2}\\
&+\frac{3\left(\sqrt{290}-17\right)}{1+\left(\sqrt{290}-17\right)^2}+\frac{3\left(\sqrt{442}-21\right)}{1+\left(\sqrt{442}-21\right)^2}+\frac{3\left(\sqrt{485}-22\right)}{1+\left(\sqrt{485}-22\right)^2}\\
&+\frac{3\left(\sqrt{730}-27\right)}{1+\left(\sqrt{730}-27\right)^2}+\frac{6\left(\sqrt{901}-30\right)}{1+\left(\sqrt{901}-30\right)^2}+\frac{6\left(\sqrt{1522}-39\right)}{1+\left(\sqrt{1522}-39\right)^2}\\
&+\frac{6\left(\sqrt{2305}-48\right)}{1+\left(\sqrt{2305}-48\right)^2}+\frac{3\left(\sqrt{3026}-55\right)}{1+\left(\sqrt{3026}-55\right)^2}+\frac{6\left(\sqrt{3137}-56\right)}{1+\left(\sqrt{3137}-56\right)^2}\\
&+\frac{6\left(\sqrt{4762}-69\right)}{1+\left(\sqrt{4762}-69\right)^2}+\frac{3\left(\sqrt{5185}-72\right)}{1+\left(\sqrt{5185}-72\right)^2}+\frac{3\left(\sqrt{5330}-73\right)}{1+\left(\sqrt{5330}-73\right)^2}\\
&+\frac{3\left(\sqrt{6401}-80\right)}{1+\left(\sqrt{6401}-80\right)^2}+\frac{3\left(\sqrt{7922}-89\right)}{1+\left(\sqrt{7922}-89\right)^2}+\frac{6\left(\sqrt{8101}-90\right)}{1+\left(\sqrt{8101}-90\right)^2}\\
&+\frac{6\left(\sqrt{11026}-105\right)}{1+\left(\sqrt{11026}-105\right)^2}+\frac{3\left(\sqrt{12322}-111\right)}{1+\left(\sqrt{12322}-111\right)^2}+\frac{6\left(\sqrt{16901}-130\right)}{1+\left(\sqrt{16901}-130\right)^2}\\
&+\frac{3\left(\sqrt{19045}-138\right)}{1+\left(\sqrt{19045}-138\right)^2}+\frac{3\left(\sqrt{23717}-154\right)}{1+\left(\sqrt{23717}-154\right)^2}+\frac{6\left(\sqrt{29242}-171\right)}{1+\left(\sqrt{29242}-171\right)^2}\\
&+\frac{3\left(\sqrt{36482}-191\right)}{1+\left(\sqrt{36482}-191\right)^2}+\frac{3\left(\sqrt{41210}-203\right)}{1+\left(\sqrt{41210}-203\right)^2}+\frac{3\left(\sqrt{44945}-212\right)}{1+\left(\sqrt{44945}-212\right)^2}\\
&+\frac{6\left(\sqrt{63505}-252\right)}{1+\left(\sqrt{63505}-252\right)^2}+\frac{3\left(\sqrt{97970}-313\right)}{1+\left(\sqrt{97970}-313\right)^2}+\frac{3\left(\sqrt{110890}-333\right)}{1+\left(\sqrt{110890}-333\right)^2}\\
&+\frac{3\left(\sqrt{171397}-414\right)}{1+\left(\sqrt{171397}-414\right)^2}+\frac{3\left(5\sqrt{74}-43\right)}{1+\left(5\sqrt{74}-43\right)^2}\Bigg).
\end{aligned}$$

PROOF. From (1),

$$u_3 = \frac{1}{3}\sum_{(l,m,n)\in A^3}\frac{z_{l,m,n}}{1+z_{l,m,n}^2}\,, \tag{13}$$

where it can be shown that

$$z_{l,m,n} = \sqrt{k^2+1}-k\,,$$

for

$$k = k_{l,m,n} = \frac{1}{2}\left(lmn+l+m+n\right)\,. \tag{14}$$

The 35 terms in the above expression for u_3 correspond to the 35 distinct values of $z_{l,m,n}$ as (l,m,n) ranges over A^3. Of the 125 terms in (13), five correspond to fixed points, and the remaining 120 correspond to points of least period 3. Of the 40 period-3 orbits, half of them are such that l, m, and n are distinct elements of A; in such cases the distinct orbits coded by (l,m,n) and (l,n,m) satisfy $z_{l,m,n} = z_{l,n,m}$ (note that (14) is symmetric in l, m, n), thus contributing 6 identical terms in the sum (13), i.e. the terms with coefficient 6 in the above expression for u_3. The other 20 period-3 orbits, for which precisely two of l, m, n are equal, contribute 3 identical terms in the sum (13), i.e. the terms with coefficient 3 in the above expression for u_3. Thus u_3 is naturally written as a sum of $35 = 5 + 10 + 20$ terms. □

Using the exact formulae of Lemmas 2, 3, and 4 we are now able to evaluate the order-3 approximation $\mathfrak{D}_3(1/2)$ to $\mathfrak{D}(1/2)$:

LEMMA 5. *For $E = E_{\{2,4,6,8,10\}}$, the order-3 approximation $\mathfrak{D}_3(1/2)$ satisfies*

$$\mathfrak{D}_3(1/2) = 1 + \delta_1 + \delta_2 + \delta_3 < -\frac{1}{20}\,. \tag{15}$$

PROOF. Using the definitions of the δ_i, and Lemmas 2, 3, and 4, we bound[1]

$$\delta_1 = -u_1 < -954/1000\,,$$

$$\delta_2 = \frac{1}{2}u_1^2 - u_2 < -102/1000\,,$$

$$\delta_3 = u_1u_2 - u_3 - \frac{1}{6}u_1^3 < 2/1000\,,$$

therefore

$$\mathfrak{D}_3(1/2) = 1 + \delta_1 + \delta_2 + \delta_3 < -54/1000 < -1/20\,.$$

□

LEMMA 6. *The error term for the approximation of $\mathfrak{D}(1/2)$ by $\mathfrak{D}_3(1/2)$ is bounded by*

$$|\mathfrak{D}(1/2) - \mathfrak{D}_3(1/2)| < \frac{1}{20}$$

PROOF. Let $D \subset \mathbb{C}$ be the disc of radius $\varrho = 3/2$ centred at $c = 1/2$. For the alphabet $A = \{2, 4, 6, 8, 10\}$ this disc has contraction ratio $\theta = 1/3$ (the point $-1 \in \partial D$ satisfies $T_2(-1) = 1$, which has distance $1/2 = \theta\varrho$ from the centre of D, see Figure 1).

For each $n \in A = \{2, 4, 6, 8, 10\}$ the function $w_n(z) = 1/(z + n)$ has maximum modulus on $\overline{D}$ when $z = c - \varrho = -1$, in other words

$$\|w_n\|_\infty = \frac{1}{n-1}\,,$$

and therefore

$$\sum_{n\in A} \|w_n\|_\infty = 1 + \frac{1}{3} + \frac{1}{5} + \frac{1}{7} + \frac{1}{9} = \frac{563}{315}\,,$$

so

$$K = \frac{\sum_{n\in A}\|w_n\|_\infty}{\theta\sqrt{1-\theta^2}} = \frac{563}{70\sqrt{2}} < 6\,.$$

Now

$$|\delta_n| \le K^n E_n(\theta) < 6^n E_n(1/3) = \frac{6^n 3^{-n(n+1)/2}}{\prod_{i=1}^n (1 - 3^{-i})} =: F_n\,,$$

[1]These bounds are conveniently checked by numerically evaluating the explicit formulae for u_1, u_2, u_3 given in Lemmas 2, 3, and 4, using either a pocket calculator or a package such as Mathematica. One finds that $u_1 = 0.95459995\ldots$, $u_2 = 0.55800098\ldots$, and $u_3 = 0.38595811\ldots$.

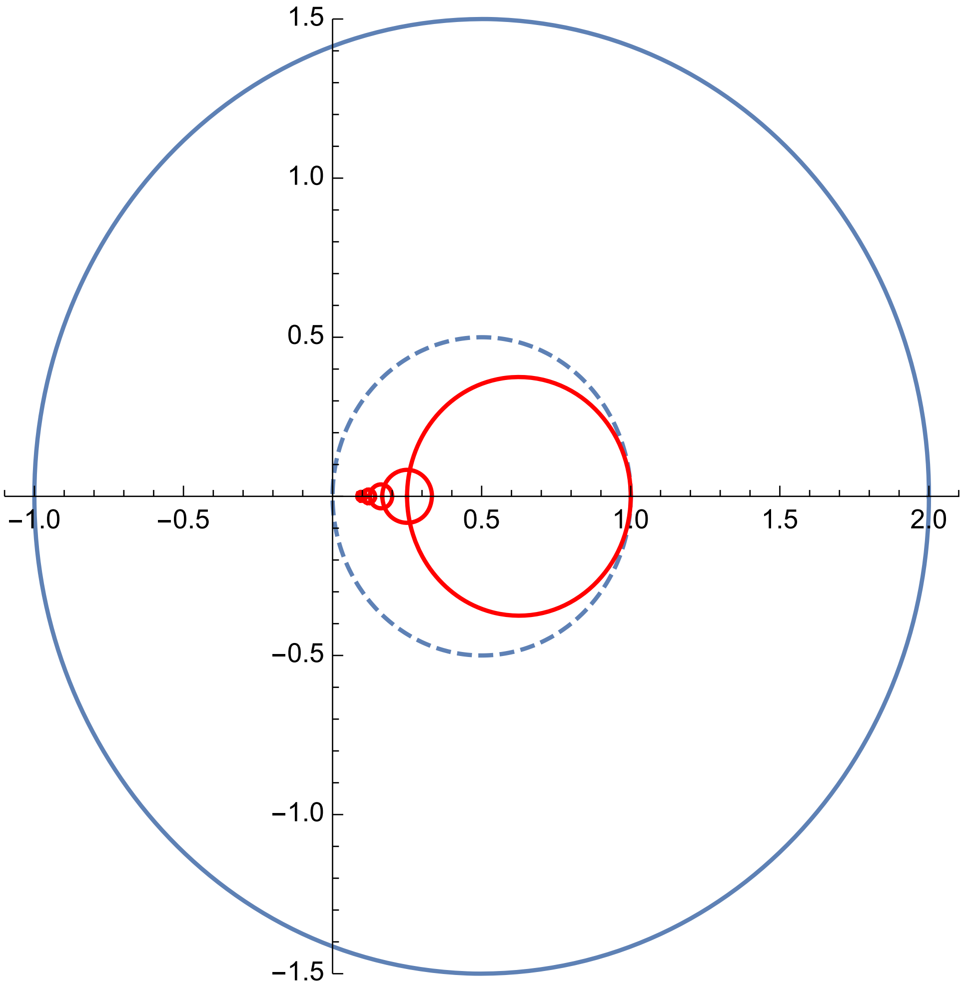


FIGURE 1. Inner disc D' (dashed) contains images $T_2(D)$, $T_4(D)$, $T_6(D)$, $T_8(D)$, $T_{10}(D)$ of the outer disc D (centre 1/2, radius 3/2), in the proof that $\dim(E_{\{2,4,6,8,10\}}) > 1/2$

from which we readily derive[2] the required bound

$$|\mathfrak{D}(1/2) - \mathfrak{D}_3(1/2)| \le \sum_{n=4}^{\infty} |\delta_n| < \sum_{n=4}^{\infty} F_n < 1/20\,.$$

□

REMARK 2. The specific disc D used in the proof of Lemma 6 ensures that error bounds are both reasonably sharp and take a conveniently simple form. Its contraction ratio 1/3 is fairly close to optimal, though in fact the minimum possible contraction ratio is slightly smaller than 3/10, and if we were wishing to establish

[2] Note that $F_4 = 81/2080 \approx 0.0389$, $F_5 = 243/251680 \approx 0.000965512$, $F_6 = 729/91611520 \approx 0.0000079$, etc., and in fact $\sum_{n=4}^{\infty} F_n = 0.039915\ldots$.

n	s_n
1	0.48423601174084654015914428125801664463082136184352
2	0.51785646889922347669500264756892828759037033720127
3	0.51735835552554373712759333961028665424316762904677
4	0.51735703035327082175724494790903719578904071340121
5	0.51735703093697422452618598486769311779169231777479
6	0.51735703093701730520259909968044128779914246471704
7	0.51735703093701730466662960310305782115301520544050
8	0.51735703093701730466662847483603679980173115413977
9	0.51735703093701730466662847483643973376603352818029
10	0.51735703093701730466662847483643973379049172430329

TABLE 1. Approximations s_n to $\dim(E_{\{2,4,6,8,10\}})$

very high accuracy rigorous bounds on $\dim(E_{\{2,4,6,8,10\}})$ then it would be preferable to work with a disc whose contraction ratio is (close to) optimal. Note that our choice of D here is not available in the case of alphabets A containing the number 1, since the point -1 on the boundary of D is then a pole for the function $1/(z+1)$ which arises in defining the associated transfer operator.

We can now prove the desired result:

THEOREM 7. *The Hausdorff dimension of $E_{\{2,4,6,8,10\}}$ is strictly larger than $1/2$.*

PROOF. For $A = \{2, 4, 6, 8, 10\}$, Lemmas 5 and 6 together give $\mathfrak{D}(1/2) < 0$ where $\mathfrak{D} = \mathfrak{D}_A$. Proposition 1 then implies that $\dim(E_A) > 1/2$. □

To end this section we provide (see Table 1) a sequence of approximations to the dimension of $E_{\{2,4,6,8,10\}}$, indicating that

$$\dim(E_{\{2,4,6,8,10\}}) = 0.517357030937017304666628474836439733 7\ldots.$$

With extra work, the majority of these empirically observed decimal digits could be rigorously justified using the techniques involving computed bounds (along the lines of §5 and §6), though in this section our preference was to establish, in a rather explicit way not relying on computer assistance, the more conservative lower bound $\dim(E_{\{2,4,6,8,10\}}) > 1/2$ which is of specific number-theoretic interest (see [**5**, p. 139] and [**23**, Lem. 2.20]).

5. The Hausdorff dimension of $E_{\{1,2,3,4,5\}}$

Here we consider the set $E_{\{1,2,3,4,5\}}$, corresponding to the choice $A = \{1, 2, 3, 4, 5\}$. The approximation s_N to $\dim(E_{\{1,2,3,4,5\}})$, based on periodic points of period up to N, is the zero (in the interval $(0, 1)$) of the function $\mathfrak{D}_N$ defined by (4); these approximations are tabulated in Table 2 for $1 \le n \le 8$. We note that the 7th and 8th approximations to $\dim(E_{\{1,2,3,4,5\}})$ share the first 13 decimal digits[3] 0.8368294436812.

[3] Note that Hensley [**17**, p. 16] gives the ten decimal digit approximation 0.8368294437, where the first 9 digits are correct, and the final digit is rounded up.

n	s_n
1	0.705879459442766674905124438813
2	0.848104427201487198901594372491
3	0.837214988477016376170810547613
4	0.836824477038318042493697933421
5	0.836829420428362177143803729319
6	0.836829443722239849891499678185
7	0.836829443681235947667216097180
8	0.836829443681208815677961682649

TABLE 2. Approximations s_n to $\dim(E_{\{1,2,3,4,5\}})$

It turns out that we can *rigorously* justify 8 of these decimal digits. Define

$$s^- = 0.83682944$$

and

$$s^+ = 0.83682945 = s^- + 10^{-8}\,.$$

We then claim:

THEOREM 8. *The Hausdorff dimension* $\dim(E_{\{1,2,3,4,5\}})$ *lies in the interval* (s^-, s^+).

PROOF. Since E_A is a subset of $\mathbb{R}$, its Hausdorff dimension is smaller than 1, and by Proposition 1 we know that $\dim(E_A)$ is the largest real zero of $\mathfrak{D}$. Our strategy is to firstly show that $\mathfrak{D}(s^-) < 0 < \mathfrak{D}(s^+)$, so that the continuous function $\mathfrak{D}$ has a zero in (s^-, s^+), and secondly show that $\mathfrak{D}$ is strictly increasing on $(s^+, 1)$, from which it follows that $\mathfrak{D}$ has no real zeros larger than s^+, hence that $\dim(E_A)$ must lie between s^- and s^+.

Let $D \subset \mathbb{C}$ be the open disc centred at c, of radius ϱ, where c is the largest real root of the polynomial

$$5c^7 + 60c^6 + 243c^5 + 309c^4 - 225c^3 - 459c^2 + 225c - 21\,,$$

so that

$$c \approx 0.871259267043988728104853432066954096301642480251564013290706298815\,,$$

and

$$\varrho = -2 + \sqrt{c^2 + 6c + 8 - 3/c}\,, \tag{16}$$

so that

$$\varrho \approx 1.24705349298248245984837857517910962469791117416655000430012735\,.$$

The relation (16) ensures that $T_1(c-\varrho)$ and $T_5(c+\varrho)$ are equidistant from c, and this common distance is denoted by $\varrho' = T_1(c-\varrho) - c = c - T_5(c+\varrho)$, so that

$$\varrho' \approx 0.730776538381714937358210535581775862495407050089163969996563349\,.$$

The specific choice of c is to ensure that the contraction ratio $\theta = \varrho'/\varrho$ is minimised, taking the value

$$\theta = \frac{\varrho'}{\varrho} \approx 0.586002559227810334771610807887260173705711718278460922051957\,.$$

Having computed the points of period up to $P = 8$ we can form the functions $s \mapsto \delta_n(s)$ for $1 \le n \le 8$, and evaluate these at $s = s^-$ to give

(17)
$$\mathfrak{D}_8(s^-) = 1 + \sum_{n=1}^{8} \delta_n(s^-) = (-7.23265042091732132359\ldots)\times 10^{-9} < -7\times 10^{-9} < 0\,,$$

and at $s = s^+$ to give

$$\mathfrak{D}_8(s^+) = 1 + \sum_{n=1}^{8} \delta_n(s^+) = (1.24148369391570553114\ldots) \times 10^{-8} > 10^{-8} > 0\,. \tag{18}$$

We now aim to show that the approximation $\mathfrak{D}_8$ is close enough to $\mathfrak{D}$ for (17) and (18) to imply, respectively, the negativity of $\mathfrak{D}(s^-)$ and the positivity of $\mathfrak{D}(s^+)$. In other words, we seek to bound the tail $\sum_{n=9}^{\infty} \delta_n(s)$, and this will be achieved by bounding the individual Taylor coefficients $\delta_n(s)$, for $n \ge 9 = P + 1$. It will turn out that for $n \ge 13$ the cruder Euler bound on $\delta_n(s)$ is sufficient, while for $P+1 = 9 \le n \le 12 = Q$ we will use the upper computed Taylor bound (cf. Remark 1) $\beta^{M,+}_{n,N,+}(s)$ for suitable $M, N \in \mathbb{N}$.

Henceforth let $Q = 12$, $M = 150$, $N = 200$, and consider the case $s = s^-$. We first evaluate the $H^2(D)$ norms of the monomial images $\mathcal{L}_{A,s}(m_k)$ for $0 \le k \le N = 200$, as

$$\begin{aligned}
\|\mathcal{L}_{A,s}(m_0)\| &= 1.18094153698482882249447608084779380079799521014296\ldots \\
\|\mathcal{L}_{A,s}(m_1)\| &= 0.50373481635455365839901987777081994881907010494221\ldots \\
\|\mathcal{L}_{A,s}(m_2)\| &= 0.25538908510961660244036590250705094646855677581007\ldots \\
&\vdots \\
\|\mathcal{L}_{A,s}(m_{200})\| &= (9.2211490601699406685842370009793893017\ldots) \times 10^{-48}.
\end{aligned}$$

Using these norms $\|\mathcal{L}_{A,s}(m_k)\|$ we then evaluate, for $1 \le n \le M = 150$, the terms $\alpha_{n,N,+}(s) = \alpha_{n,200,+}(s)$ defined (cf. (9)) by

$$\alpha_{n,N,+}(s) = \left(\sum_{k=n-1}^{N} \|\mathcal{L}_{A,s}(m_k)\|^2 + \left(\sum_{i=1}^{5} \|w_{i,s}\|_\infty \right)^2 \frac{\theta^{2(N+1)}}{1-\theta^2} \right)^{1/2}$$

to be

$$\begin{aligned}
\alpha_{1,200,+}(s) &= 1.31924766289256695924356827596610055341618618514631\ldots \\
\alpha_{2,200,+}(s) &= 0.58804037469497804159060266597641325581551232133109\ldots \\
\alpha_{3,200,+}(s) &= 0.30338542658416252872670480452558662518433118485741\ldots \\
&\vdots \\
\alpha_{150,200,+}(s) &= (8.4073197947570136649265418048602686584245204793167\ldots) \\
&\quad \times 10^{-36}.
\end{aligned}$$

The terms $\alpha_{n,200,+}(s)$ are then used to form the upper computed Taylor bounds $\beta^{M,+}_{n,N,+}(s) = \beta^{M,-}_{n,N,+}(s) + \sum_{l=0}^{n-1} J^{n-l}_{Q,N,s}\, \beta^{M,-}_{l,N,+}(s)\, \theta^{M(n-l)} E_{n-l}(\theta)$, where

$$\beta^{M,-}_{n,N,+}(s) = \beta^{150,-}_{n,200,+}(s) = \sum_{i_1 < \ldots < i_n \le 150} \prod_{j=1}^{n} \alpha_{i_j,200,+}(s)\,,$$

which for $9 \leq n \leq 12 = Q$ are[4]

$$\beta^{M,+}_{9,N,+}(s) = (3.8691484792014233501009508862660178562663259339\,93\ldots) \times 10^{-9}$$

$$\beta^{M,+}_{10,N,+}(s) = (2.0410281556300938956257995289307647109627120034\,14\ldots) \times 10^{-11}$$

$$\beta^{M,+}_{11,N,+}(s) = (6.1309246226139368378720041951472354024865024502\,29\ldots) \times 10^{-14}$$

$$\beta^{M,+}_{12,N,+}(s) = (1.0522363626350277460656303574730842052357778811\,099\ldots) \times 10^{-16}$$

so in particular

$$\sum_{n=9}^{12} |\delta_n(s)| \leq \sum_{n=9}^{12} \beta^{M,+}_{n,N,+}(s) < 3.9 \times 10^{-9}. \tag{19}$$

It remains to derive the Euler bounds on the Taylor coefficients $\delta_n(s)$ for $n \geq 13$. For $s > 0$ and $i \in \{1,2,3,4,5\}$, the function $w_{i,s}(z) = 1/(z+i)^{2s}$ has maximum modulus on D when $z = c - \varrho$, so

$$\|w_{i,s}\|_\infty = 1/(i + c - \varrho)^{2s}\,. \tag{20}$$

A computation using (20) gives

$$\|w_{1,s}\|_\infty \leq 2.2006525312032484040444791042266424054625533414310150581771 55\,,$$

$$\|w_{2,s}\|_\infty \leq 0.4440774658899549894209825596616278157148192709610040729216 69\,,$$

$$\|w_{3,s}\|_\infty \leq 0.1989484070468766242919273349564959863224875881198236032001 26\,,$$

$$\|w_{4,s}\|_\infty \leq 0.1158960010977102308020238251807916905536186118173927712063 40\,,$$

$$\|w_{5,s}\|_\infty \leq 0.0770823001494264304016599133908923697838630633552897879251 34\,,$$

thus

$$\sum_{i=1}^{5} \|w_{i,s}\|_\infty \leq 3.0366567053872166789610727374164502678373418756845252934 30\,,$$

and therefore

$$K_s = \frac{\sum_{i=1}^{5} \|w_{i,s}\|_\infty}{\theta\sqrt{1-\theta^2}} \leq 6.39507165244054791777743776407948610 7\,.$$

Now $|\delta_n(s)| \leq K_s^n E_n(\theta)$, and we readily compute that

$$K_s^{13} E_{13}(\theta) < (1.4001102011420297343801031463546031641312628016 5\ldots) \times 10^{-10}\,,$$

$$K_s^{14} E_{14}(\theta) < (5.0448172369716376790742252310568321303894463405 4\ldots) \times 10^{-13}\,,$$

and the super-exponential decay of the terms $K_s^n E_n(\theta)$ means we easily bound

$$\left| \sum_{n=13}^{\infty} \delta_n(s) \right| \leq \sum_{n=13}^{\infty} K_s^n E_n(\theta) < 1.5 \times 10^{-10}\,. \tag{21}$$

Combining (21) with (19) gives, for $s = s^-$,

$$\left| \sum_{n=9}^{\infty} \delta_n(s) \right| < 4 \times 10^{-9}\,. \tag{22}$$

[4]Although not needed in this proof, we record here that the values of $\beta^{M,+}_{n,N,+}(s)$ for $1 \leq n \leq 8$ are $\beta^{M,+}_{1,N,+}(s) \approx 2.58$, $\beta^{M,+}_{2,N,+}(s) \approx 2.22$, $\beta^{M,+}_{3,N,+}(s) \approx 0.84$, $\beta^{M,+}_{4,N,+}(s) \approx 0.16$, $\beta^{M,+}_{5,N,+}(s) \approx 0.015$, $\beta^{M,+}_{6,N,+}(s) \approx 0.00085$, $\beta^{M,+}_{7,N,+}(s) \approx 0.000025$, $\beta^{M,+}_{8,N,+}(s) \approx 4.15 \times 10^{-7}$.

Combining (22) with (17) then gives

$$\mathfrak{D}(s^-) = 1 + \sum_{n=1}^{\infty} \delta_n(s^-) < -3 \times 10^{-9} < 0\,. \tag{23}$$

We now show that $\mathfrak{D}(s^+)$ is positive. In view of (18), for this it is sufficient to show that $|\sum_{n=9}^{\infty} \delta_n(s)| < 10^{-8}$ for $s = s^+$. In fact the stronger inequality (22) (which we have proved for $s = s^-$) can also be established for $s = s^+$, using the same general method as for $s = s^-$, since the intermediate computed values for the norms $\|\mathcal{L}_{A,s}(m_k)\|$, the terms $\alpha_{n,N,+}(s)$, the computed Taylor bounds $\beta_{n,N,+}^{M,+}(s)$, and the Euler bounds $K_s^n E_n(\theta)$, are sufficiently close to those for $s = s^- = s^+ - 10^{-8}$. Combining (18) with inequality (22) for $s = s^+$ gives the required positivity

$$\mathfrak{D}(s^+) = 1 + \sum_{n=1}^{\infty} \delta_n(s^+) > 0\,. \tag{24}$$

Since $\mathfrak{D}$ is continuous, (23) and (24) imply that it has a zero in (s^-, s^+), and in particular that $\dim(E_A)$, as the largest zero of $\mathfrak{D}$ (by Proposition 1), is larger than s^-. To prove that $\dim(E_A) < s^+$ it now suffices to show that $\mathfrak{D}$ is strictly increasing on $(s^+, 1)$, and hence has no zeros in this interval. For this we use that the function $\mathfrak{D}_8(s) = 1 + \sum_{n=1}^{8} \delta_n(s)$ is available to us in closed form, together with an estimate on the derivative of the remainder function

$$\mathfrak{R}_8(s) := \mathfrak{D}(s) - \mathfrak{D}_8(s) = \sum_{n=9}^{\infty} \delta_n(s)\,.$$

In particular, $\mathfrak{D}_8$ can be shown to be both strictly increasing and strictly concave on the interval $(s^+, 1)$ (cf. Figure 2, showing the restriction of $\mathfrak{D}_8$ to $[0, 1]$), with

$$\mathfrak{D}_8'(s) > \mathfrak{D}_8'(1) = 1.3546901785\ldots > \frac{13}{10} \quad \text{for all } s \in (s^+, 1)\,. \tag{25}$$

Let U denote the ε-neighbourhood in $\mathbb{C}$ of the real interval $(s^+, 1)$, where for concreteness we choose $\varepsilon = 1/10$. We shall bound the modulus of $\mathfrak{R}_8$ on U via Euler bounds on the coefficients $\delta_n(s)$ for $n > 8$, $s \in U$, and then use Cauchy's integral formula to derive a bound on $\mathfrak{R}_8'(s)$ for $s \in (s^+, 1)$. Recall (see Theorem 6) that $|\delta_n(s)| \le K_s^n E_n(\theta)$, where $E_n(\theta) = \theta^{n(n+1)/2} \prod_{i=1}^{n} (1 - \theta^i)^{-1}$ is independent of s, and

$$K_s = \frac{\sum_{i=1}^{5} \|w_{i,s}\|_\infty}{\theta\sqrt{1-\theta^2}} = \frac{\sum_{i=1}^{5} (i + c - \varrho)^{-2s}}{\theta\sqrt{1-\theta^2}}\,.$$

It is readily shown that

$$\sup_{s \in U} |K_s| = K_{1+\varepsilon} = K_{11/10} = 7.1122943065860651834 8\ldots\,,$$

so that

$$\sup_{s \in U} |\delta_9(s)| \le K_{11/10}^9 E_9(\theta) = 0.0102436709523374009 2\ldots\,,$$

$$\sup_{s \in U} |\delta_{10}(s)| \le K_{11/10}^{10} E_{10}(\theta) = 0.00034957413642133622\ldots\,,$$

$$\sup_{s \in U} |\delta_{11}(s)| \le K_{11/10}^{11} E_{11}(\theta) = 0.00000697687020201114\ldots\,,$$

$$\sup_{s \in U} |\delta_{12}(s)| \le K_{11/10}^{12} E_{12}(\theta) = 0.00000008150368808892\ldots\,,$$

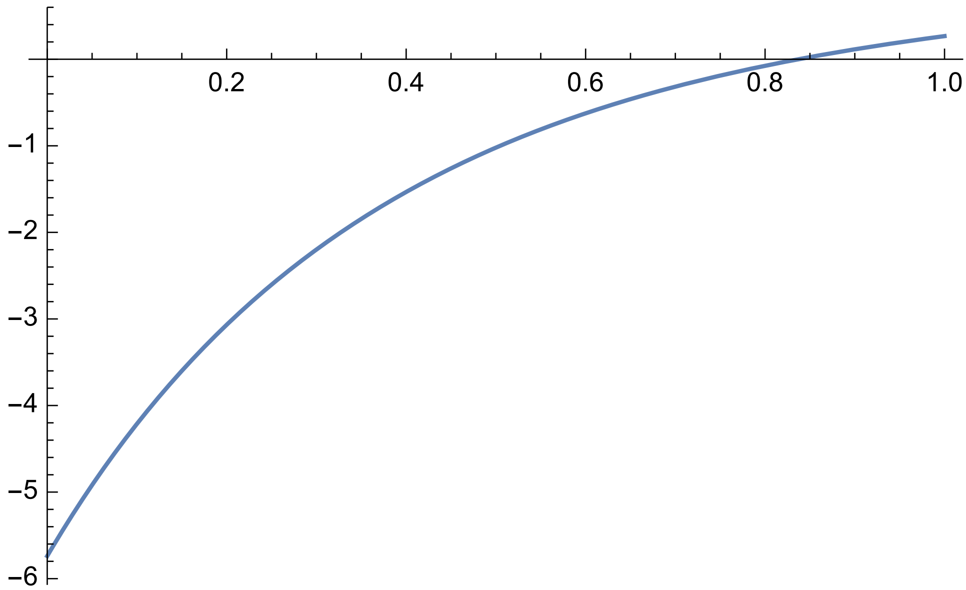


FIGURE 2. Order-8 approximation $\mathfrak{D}_8$ to the dimension determinant for $E_{\{1,2,3,4,5\}}$.

and we can therefore bound

$$\sup_{s\in U}|\mathfrak{R}_8(s)| \le \sum_{n=9}^{\infty}\sup_{s\in U}|\delta_n(s)| < \frac{11}{1000}\,. \tag{26}$$

If C_s denotes the positively oriented circular contour of radius $\varepsilon = 1/10$, centred at s, then Cauchy's integral formula gives the derivative formula $\mathfrak{R}_8'(s) = \frac{1}{2\pi i}\oint_{C_s}\frac{\mathfrak{R}_8(t)}{(t-s)^2}\,dt$, and U is the union over $s\in(s^+,1)$ of the open discs bounded by the C_s, so (26) yields

$$|\mathfrak{R}_8'(s)| \le \frac{1}{\varepsilon^2}\sup_{t\in C_s}|\mathfrak{R}_8(t)| \le 100\sup_{t\in U}|\mathfrak{R}_8(t)| < \frac{11}{10} \quad \text{for all } s\in(s^+,1)\,.$$

In particular,

$$\mathfrak{R}_8'(s) > -\frac{11}{10} \quad \text{for all } s\in(s^+,1)\,, \tag{27}$$

so combining (25) and (27) gives

$$\mathfrak{D}'(s) = \mathfrak{D}_8'(s) + \mathfrak{R}_8'(s) > 0 \quad \text{for all } s\in(s^+,1)\,,$$

so indeed $\mathfrak{D}$ is strictly increasing on $(s^+,1)$, as required. □

REMARK 3. The analysis of Bourgain-Kontorovich and Huang also applies to more general finite subsets $A\subset\mathbb{N}$, see [**18**]. More precisely, if $\dim(E_A) > \frac{5}{6}$ then the corresponding subset $Q_A\subset\mathbb{N}$ has density one. Therefore, it is natural to consider other non-sequential finite subsets A for which we can rigorously show $\dim(E_A) > \frac{5}{6}$.

n	s_n
1	0.742538972647559226233764770933
2	0.878373312250454078800953132613
3	0.867983314494266000322362181011
4	0.867614537223019698282665367406
5	0.867619151810964612388367917252
6	0.867619173277394444047871397558
7	0.867619173240135215103752503105
8	0.867619173240110928010919906321

TABLE 3. Approximations s_n to $\dim(E_{\{1,2,3,4,5,6\}})$

Using the same method as that used in this section, one can show rigorously that the dimensions of the sets E_A associated to the choices $A = \{1, 2, 3, 5, 6, 8\}$ or $A = \{1, 2, 3, 4, 6, 18\}$, for example, are greater than $\frac{5}{6}$.

6. The Hausdorff dimension of $E_{\{1,2,3,4,5,6\}}$

Here we consider the set $E_{\{1,2,3,4,5,6\}}$, corresponding to the choice $A = \{1, 2, 3, 4, 5, 6\}$. The approximation s_N to $\dim(E_{\{1,2,3,4,5,6\}})$, based on periodic points of period up to N, is the zero (in the interval $(0, 1)$) of the function $\mathfrak{D}_N$ defined by (4); these approximations are tabulated in Table 3 for $1 \leq n \leq 8$. We note that the 7th and 8th approximations to $\dim(E_{\{1,2,3,4,5,6\}})$ share the first 13 decimal digits 0.8676191732401.

It turns out that we can *rigorously* justify 7 of these decimal digits. Define

$$s^- = 0.8676191$$

and

$$s^+ = 0.8676192 = s^- + 10^{-7}\,.$$

We then claim:

THEOREM 9. *The Hausdorff dimension* $\dim(E_{\{1,2,3,4,5,6\}})$ *lies in the interval* (s^-, s^+).

PROOF. The strategy of proof is similar to that used for Theorem 8, firstly showing that $\mathfrak{D}(s^-) < 0 < \mathfrak{D}(s^+)$ so that $\mathfrak{D}$ has a zero in (s^-, s^+), and secondly arguing that this is the largest zero of $\mathfrak{D}$, hence must be $\dim(E_A)$.

Let $D \subset \mathbb{C}$ be the open disc centred at c, of radius ϱ, where this time

$$c \approx 0.888786621704996501948480357049568065602437524401186717911139539201$$

is chosen as the largest real root of the polynomial

$$384c^7 + 5376c^6 + 25872c^5 + 42560c^4 - 16660c^3 - 67228c^2 + 26803c - 2744\,,$$

and

$$\begin{aligned}\varrho &= -\frac{5}{2} + \frac{1}{2}\sqrt{4c^2 + 28c + 45 - 14/c} \\ &\approx 1.284639341742533191143484074021163452454469\,.\end{aligned}$$

It follows that

$$\begin{aligned}\varrho' &= T_1(c-\varrho) - c = c - T_6(c+\varrho)\\ &\approx 0.76643890552427727077005511585427320750401107808160\,,\end{aligned}$$

and therefore

$$\theta = \frac{\varrho'}{\varrho} \approx 0.59661796164879293682899603757410251596387247454335884230857 3\,.$$

The points of period up to $P = 8$ determine the functions $s \mapsto \delta_n(s)$ for $1 \le n \le 8$, which when evaluated at $s = s^-$ and $s = s^+$ give

(28)

$$\mathfrak{D}_8(s^-) = 1 + \sum_{n=1}^{8} \delta_n(s^-) = (-1.498373759369204270864\ldots) \times 10^{-7} < -10^{-7} < 0\,,$$

(29)

$$\mathfrak{D}_8(s^+) = 1 + \sum_{n=1}^{8} \delta_n(s^+) = (5.474638165609240513579\ldots) \times 10^{-8} > 5 \times 10^{-8} > 0\,.$$

We now claim that $\mathfrak{D}_8$ is close enough to $\mathfrak{D}$ for the inequalities (28) and (29) to imply that $\mathfrak{D}(s^-) < 0 < \mathfrak{D}(s^+)$, and will establish this by bounding $\delta_n(s)$, for $n \ge 9 = P + 1$. As previously, for $n \ge 13$ the Euler bound on $\delta_n(s)$ turns out to be sufficient, while for $9 \le n \le 12 =: Q$ we use upper computed Taylor bounds $\beta^{M,+}_{n,N,+}(s)$, where once again we set $M := 150$, $N := 200$. To introduce some variety in the part of the proof presented in full, and in recognition of the fact that in the present case $\mathfrak{D}_8(s^+)$ is closer to zero than $\mathfrak{D}_8(s^-)$ is, we here consider the case $s = s^+$.

The norms $\|\mathcal{L}_{A,s}(m_k)\|$ are computed via numerical integration, and then used to form the terms $\alpha_{n,200,+}(s)$, which are then used to form the upper computed Taylor bounds which for $9 \le n \le 12$ take the values

$$\beta^{M,+}_{9,N,+}(s) = (1.262194624669540668569841998650141041089448447560 1\ldots) \times 10^{-8}$$

$$\beta^{M,+}_{10,N,+}(s) = (8.314966430413518627081622024687687663503710477628\ldots) \times 10^{-11}$$

$$\beta^{M,+}_{11,N,+}(s) = (3.176610018228192242136810148998407171840692198466\ldots) \times 10^{-13}$$

$$\beta^{M,+}_{12,N,+}(s) = (7.061524069747938884792482724386219757269895805839\ldots) \times 10^{-16}\,,$$

from which

$$\sum_{n=9}^{12} |\delta_n(s)| \le \sum_{n=9}^{12} \beta^{M,+}_{n,N,+}(s) < 1.3 \times 10^{-8}. \tag{30}$$

To compute the Euler bounds on $\delta_n(s)$ for $n \ge 13$ we note, as previously, that $\|w_{i,s}\|_\infty = 1/(i + c - \varrho)^{2s}$, whence $\|w_{1,s}\|_\infty = 2.39\ldots$, $\|w_{2,s}\|_\infty = 0.44\ldots$, $\|w_{3,s}\|_\infty = 0.18\ldots$, $\|w_{4,s}\|_\infty = 0.10\ldots$, $\|w_{5,s}\|_\infty = 0.07\ldots$, $\|w_{6,s}\|_\infty = 0.05\ldots$, and

$$\sum_{i=1}^{6} \|w_{i,s}\|_\infty \le 3.256977065218374220933841250657 77\,,$$

therefore

$$K_s = \frac{\sum_{i=1}^{6} \|w_{i,s}\|_\infty}{\theta\sqrt{1-\theta^2}} \le 6.802359696999181386288200501725510191455\,.$$

It follows that[5]

$$K_s^{13}E_{13}(\theta) < (1.751608670306048305544710571625984526147775\ldots)\times 10^{-9}\,,$$
$$K_s^{14}E_{14}(\theta) < (8.632960731433444691012413481027827464799512\ldots)\times 10^{-12}\,,$$

and

$$\left|\sum_{n=13}^{\infty}\delta_n(s)\right| \le \sum_{n=13}^{\infty} K_s^n E_n(\theta) < 2\times 10^{-9}\,, \tag{31}$$

so (30), (31) together give, for $s=s^+$,

$$\left|\sum_{n=9}^{\infty}\delta_n(s)\right| < 1.5\times 10^{-8}\,, \tag{32}$$

and combining (32) with (29) gives

$$\mathfrak{D}(s^+) = 1+\sum_{n=1}^{\infty}\delta_n(s^+) > 3.5\times 10^{-8} > 0\,. \tag{33}$$

It remains to show that $\mathfrak{D}(s^-)$ is negative. In view of (28), for this it is sufficient to show that $|\sum_{n=9}^{\infty}\delta_n(s)| < 10^{-7}$ for $s=s^-$. In fact the stronger inequality (32) (which we have proved for $s=s^+$) can also be established for $s=s^-$, using the same general method as for $s=s^+$, since the intermediate computed values for $\|\mathcal{L}_{A,s}(m_k)\|$, $\alpha_{n,N,+}(s)$, $\beta_{n,N,+}^{M,+}(s)$, and $K_s^n E_n(\theta)$, are sufficiently close to those for $s=s^+=s^-+10^{-8}$. Combining (28) with inequality (32) for $s=s^-$ gives the required negativity

$$\mathfrak{D}(s^-) = 1+\sum_{n=1}^{\infty}\delta_n(s^-) < 0\,. \tag{34}$$

Since $\mathfrak{D}$ is continuous, (33) and (34) imply that it has a zero in (s^-,s^+), and in particular that $\dim(E_A)$, as the largest zero of $\mathfrak{D}$ (by Proposition 1), is larger than s^-.

To prove that $\dim(E_A) < s^+$ it now suffices to show that $\mathfrak{D}$ has no zeros in $(s^+,1)$. For this, it is technically convenient to deviate slightly from the approach used in the proof of Theorem 8, by firstly establishing that $\mathfrak{D}$ is strictly increasing on $(s^+,9/10)$ (hence has no zeros in this interval, since $\mathfrak{D}(s^+)>0$), and then showing directly that $\mathfrak{D}$ is strictly positive on $[9/10,1]$.

The function $\mathfrak{D}_8$ can be shown to be both strictly increasing and strictly concave on the interval $(s^+,9/10)$, with

$$\mathfrak{D}_8'(s) > \mathfrak{D}_8'(9/10) = 1.8898838248\ldots > \frac{3}{2} \quad \text{for all } s\in(s^+,9/10)\,. \tag{35}$$

Define $\mathfrak{R}_8(s) := \mathfrak{D}(s)-\mathfrak{D}_8(s) = \sum_{n=9}^{\infty}\delta_n(s)$, and let U denote the ε-neighbourhood in $\mathbb{C}$ of the interval $(s^+,9/10)$, where $\varepsilon=1/5$. Now $|\delta_n(s)|\le K_s^n E_n(\theta)$ (by Theorem 6), where

$$K_s = \frac{\sum_{i=1}^{6}\|w_{i,s}\|_\infty}{\theta\sqrt{1-\theta^2}} = \frac{\sum_{i=1}^{6}(i+c-\varrho)^{-2s}}{\theta\sqrt{1-\theta^2}}\,,$$

[5]Note that $K_s^{12}E_{12}(\theta) = (2.119\ldots)\times 10^{-7}$, which is slightly too large for our purposes, thus justifying the choice of $Q=12$ as the largest index for which the computed Taylor bound, rather than the Euler bound, is used.

and it is readily shown that

$$\sup_{s\in U} |K_s| = K_{9/10+\varepsilon} = K_{11/10} = 7.56580745219371800335\dots ,$$

so that

$$\sup_{s\in U} |\delta_9(s)| \le K_{11/10}^9 E_9(\theta) = 0.04376686701280541755\dots ,$$

$$\sup_{s\in U} |\delta_{10}(s)| \le K_{11/10}^{10} E_{10}(\theta) = 0.00190306136091161412\dots ,$$

$$\sup_{s\in U} |\delta_{11}(s)| \le K_{11/10}^{11} E_{11}(\theta) = 0.00004925505297772416\dots ,$$

$$\sup_{s\in U} |\delta_{12}(s)| \le K_{11/10}^{12} E_{12}(\theta) = 0.00000075953226513723\dots ,$$

and we can therefore bound

$$\sup_{s\in U} |\mathfrak{R}_8(s)| \le \sum_{n=9}^{\infty} \sup_{s\in U} |\delta_n(s)| < \frac{1}{20}\,. \tag{36}$$

If C_s denotes the positively oriented circular contour of radius $\varepsilon = 1/5$, centred at s, then Cauchy's integral formula together with (36) yields

$$|\mathfrak{R}_8'(s)| \le \frac{1}{\varepsilon^2} \sup_{t\in C_s} |\mathfrak{R}_8(t)| \le 25 \sup_{t\in U} |\mathfrak{R}_8(t)| < \frac{5}{4} \quad \text{for all } s \in (s^+, 9/10)\,,$$

and in particular,

$$\mathfrak{R}_8'(s) > -\frac{5}{4} \quad \text{for all } s \in (s^+, 9/10)\,, \tag{37}$$

so combining (35) and (37) gives

$$\mathfrak{D}'(s) = \mathfrak{D}_8'(s) + \mathfrak{R}_8'(s) > 0 \quad \text{for all } s \in (s^+, 9/10)\,,$$

so indeed $\mathfrak{D}$ is strictly increasing on $(s^+, 9/10)$, as claimed.

It remains to show that $\mathfrak{D}$ has no zeros in the interval $[9/10, 1]$. Since $\mathfrak{D}_8$ is increasing on this interval,

$$\mathfrak{D}_8(s) \ge \mathfrak{D}_8(9/10) = 0.06368315529812853238\dots > \frac{1}{20} \quad \text{for all } s \in [9/10, 1]\,. \tag{38}$$

Now $s \mapsto K_s^n$ is increasing on $[9/10, 1]$, so if $s \in [9/10, 1]$ then

$$|\delta_n(s)| \le K_s^n E_n(\theta) \le K_1^n E_n(\theta) = \frac{K_1^n \theta^{n(n+1)/2}}{\prod_{i=1}^n (1-\theta^i)} < A K_1^n \theta^{n(n+1)/2}$$

where $A := \prod_{i=1}^n (1-\theta^i)^{-1} = 6.780731869\dots$, therefore

$$\sum_{n=9}^{\infty} |\delta_n(s)| < A \sum_{n=9}^{\infty} K_1^n \theta^{n(n+1)/2} < A K_1^9 \theta^{45} \sum_{i=0}^{\infty} (K_1 \theta^{10})^i = \frac{A K_1^9 \theta^{45}}{1 - K_1 \theta^{10}}$$

$$= 0.02845\dots < \frac{3}{100}\,,$$

and hence

$$\sup_{s\in[9/10,1]} |\mathfrak{R}_8(s)| \le \sup_{s\in[9/10,1]} \sum_{n=9}^{\infty} |\delta_n(s)| < \frac{3}{100}\,. \tag{39}$$

From (38) and (39) it follows that

$$\mathfrak{D}(s) = \mathfrak{D}_8(s) + \mathfrak{R}_8(s) > \frac{1}{20} - \frac{3}{100} > 0 \quad \text{for all } s \in [9/10, 1],$$

so indeed $\mathfrak{D}$ has no zeros in the interval $[9/10, 1]$, and the proof is complete. ☐

References

[1] O. F. Bandtlow, *Resolvent estimates for operators belonging to exponential classes*, Integral Equations Operator Theory **61** (2008), no. 1, 21–43, DOI 10.1007/s00020-008-1571-z. MR2414438

[2] O. F. Bandtlow and O. Jenkinson, *Explicit eigenvalue estimates for transfer operators acting on spaces of holomorphic functions*, Adv. Math. **218** (2008), no. 3, 902–925, DOI 10.1016/j.aim.2008.02.005. MR2414325

[3] T. Bedford, *Applications of dynamical systems theory to fractals—a study of cookie-cutter Cantor sets*, Fractal geometry and analysis (Montreal, PQ, 1989), NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., vol. 346, Kluwer Acad. Publ., Dordrecht, 1991, pp. 1–44, DOI 10.1098/rspa.1975.0163. MR1140719

[4] I. Borosh and H. Niederreiter, *Optimal multipliers for pseudorandom number generation by the linear congruential method*, BIT **23** (1983), no. 1, 65–74, DOI 10.1007/BF01937326. MR689604

[5] J. Bourgain and A. Kontorovich, *On Zaremba's conjecture*, Ann. of Math. (2) **180** (2014), no. 1, 137–196, DOI 10.4007/annals.2014.180.1.3. MR3194813

[6] R. Bowen, *Hausdorff dimension of quasicircles*, Inst. Hautes Études Sci. Publ. Math. **50** (1979), 11–25. MR556580

[7] R. T. Bumby, *Hausdorff dimensions of Cantor sets*, J. Reine Angew. Math. **331** (1982), 192–206, DOI 10.1515/crll.1982.331.192. MR647383

[8] R. T. Bumby, *Hausdorff dimension of sets arising in number theory*, Number theory (New York, 1983), Lecture Notes in Math., vol. 1135, Springer, Berlin, 1985, pp. 1–8, DOI 10.1007/BFb0074599. MR803348

[9] T. W. Cusick, *Continuants with bounded digits*, Mathematika **24** (1977), no. 2, 166–172, DOI 10.1112/S0025579300009050. MR0472721

[10] T. W. Cusick, *Continuants with bounded digits. II*, Mathematika **25** (1978), no. 1, 107–109, DOI 10.1112/S002557930000930X. MR0498413

[11] L. Euler, *Introductio in Analysin Infinitorum*, Marcum-Michaelem Bousquet, Lausannae, 1748.

[12] K. Falconer, *Fractal geometry*, Mathematical foundations and applications, John Wiley & Sons, Ltd., Chichester, 1990. MR1102677

[13] R. S. Falk and R. D. Nussbaum, *C^m eigenfunctions of Perron-Frobenius operators and a new approach to numerical computation of Hausdorff dimension: applications in $\mathbb{R}^1$*, J. Fractal Geom. **5** (2018), no. 3, 279–337, DOI 10.4171/JFG/62. MR3827801

[14] I. J. Good, *The fractional dimensional theory of continued fractions*, Proc. Cambridge Philos. Soc. **37** (1941), 199–228, DOI 10.1017/s030500410002171x. MR4878

[15] G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*, 5th ed., The Clarendon Press, Oxford University Press, New York, 1979. MR568909

[16] D. Hensley, *The Hausdorff dimensions of some continued fraction Cantor sets*, J. Number Theory **33** (1989), no. 2, 182–198, DOI 10.1016/0022-314X(89)90005-X. MR1034198

[17] D. Hensley, *A polynomial time algorithm for the Hausdorff dimension of continued fraction Cantor sets*, J. Number Theory **58** (1996), no. 1, 9–45, DOI 10.1006/jnth.1996.0058. MR1387719

[18] S. Huang, *An improvement to Zaremba's conjecture*, Geom. Funct. Anal. **25** (2015), no. 3, 860–914, DOI 10.1007/s00039-015-0327-6. MR3361774

[19] S. Huang, *An Improvement to Zaremba's Conjecture*, ProQuest LLC, Ann Arbor, MI, 2015. Thesis (Ph.D.)–Yale University. MR3407292

[20] O. Jenkinson, *On the density of Hausdorff dimensions of bounded type continued fraction sets: the Texan conjecture*, Stoch. Dyn. **4** (2004), no. 1, 63–76, DOI 10.1142/S0219493704000900. MR2069367

[21] O. Jenkinson and M. Pollicott, *Computing the dimension of dynamically defined sets: E_2 and bounded continued fractions*, Ergodic Theory Dynam. Systems **21** (2001), no. 5, 1429–1445, DOI 10.1017/S0143385701001687. MR1855840

[22] O. Jenkinson and M. Pollicott, *Rigorous effective bounds on the Hausdorff dimension of continued fraction Cantor sets: a hundred decimal digits for the dimension of E_2*, Adv. Math. **325** (2018), 87–115, DOI 10.1016/j.aim.2017.11.028. MR3742587

[23] A. Kontorovich, *From Apollonius to Zaremba: local-global phenomena in thin orbits*, Bull. Amer. Math. Soc. (N.S.) **50** (2013), no. 2, 187–228, DOI 10.1090/S0273-0979-2013-01402-2. MR3020826

[24] R. D. Mauldin and M. Urbański, *Graph directed Markov systems*, Cambridge Tracts in Mathematics, vol. 148, Cambridge University Press, Cambridge, 2003. Geometry and dynamics of limit sets. MR2003772

[25] D. Ruelle, *Zeta-functions for expanding maps and Anosov flows*, Invent. Math. **34** (1976), no. 3, 231–242, DOI 10.1007/BF01403069. MR0420720

[26] J. Shallit, *Real numbers with bounded partial quotients: a survey*, Enseign. Math. (2) **38** (1992), no. 1-2, 151–187. MR1175525

[27] B. Simon, *Trace ideals and their applications*, London Mathematical Society Lecture Note Series, vol. 35, Cambridge University Press, Cambridge-New York, 1979. MR541149

[28] S. K. Zaremba, *La méthode des "bons treillis" pour le calcul des intégrales multiples* (French, with English summary), Applications of number theory to numerical analysis (Proc. Sympos., Univ. Montreal, Montreal, Que., 1971), Academic Press, New York, 1972, pp. 39–119. MR0343530

School of Mathematical Sciences, Queen Mary, University of London, Mile End Road, London, E1 4NS, UK
Email address: `omj@maths.qmul.ac.uk`

Department of Mathematics, Warwick University, Coventry, CV4 7AL, UK
Email address: `masdbl@warwick.ac.uk`

Contemporary Mathematics
Volume **744**, 2020
https://doi.org/10.1090/conm/744/14981

Volume growth for infinite graphs and translation surfaces

P. Colognese and M. Pollicott

This paper is dedicated to the memory of Sergeiy Kolyada

ABSTRACT. In this note we give asymptotic estimates for the volume growth associated to suitable infinite graphs. Our main application is to give an asymptotic estimate for the volume growth associated to translation surfaces.

1. Introduction

We begin by recalling the definition of volume entropy for compact Riemannian manifolds due to Manning [**7**]. Let M be a compact manifold with Riemannian metric ρ and universal cover $\widetilde{M}$ equipped with the lifted metric $\widetilde{\rho}$. Fix a point $c \in \widetilde{M}$ and consider a ball $B(c,R)$ of radius $R > 0$ centred at c.

DEFINITION 1.1. The *volume entropy* of M is defined by

$$h = h(M,\rho) := \lim_{R\to\infty} \frac{1}{R} \log \mathrm{Vol}_{\tilde{\rho}}(B(c,R)),$$

where $\mathrm{Vol}_{\tilde{\rho}}$ denotes the Riemannian volume on $\widetilde{M}$ with respect to $\widetilde{\rho}$.

For manifolds (M,ρ) of non-positive curvature this coincides with the topological entropy h of the associated geodesic flow [**7**]. In the case of manifolds with negative sectional curvature, Margulis [**9**] showed in his thesis that there is a simple asymptotic formula: There exists $C > 0$ such that

$$\lim_{R\to+\infty} \frac{\mathrm{Vol}_{\tilde{\rho}}(B(c,R))}{e^{hR}} = C.$$

A closely related result in [**9**] gave an asymptotic formula for the number $\Pi(x,R)$ of geodesic arcs starting and finishing at a given point x of length at most R: There exists $D > 0$ such that

$$\lim_{R\to+\infty} \frac{\Pi(x,R)}{e^{hR}} = D.$$

A related notion of volume entropy was considered for directed, finite, connected, non-cyclic graphs without terminal vertices by Lim in [**6**]. In this note we extend Lim's definition of volume entropy to suitable infinite graphs and show the analogue of Margulis' result in this context (Theorem 2.1). As an application we

2010 *Mathematics Subject Classification.* Primary 05C63, 37C35; Secondary 37D40.
Key words and phrases. infinite graph, volume entropy, translation surfaces.
The second author was partially supported by EPSRC grant EP/M001903/1 and ERC grant 833802/Resonances.

show a version of Margulis' theorem for the natural analogue of volume growth for translation surfaces (Theorem 6.3). Independently, Eskin and Rafi have proved a parallel result for closed geodesics (see Remark 6.13.4).

This note originated as a summer MPhil project of the first author. It may have been possible to apply the transfer operator methods in [**15**], but instead we employ a more direct and elementary approach.

We are grateful to A. Eskin, J. Chaika, R. Sharp, S. Ghazouani and the three anonymous referees for their useful comments.

2. Infinite Graphs

In this section we will introduce the types of graphs we shall we working with as well as basic definitions which will be used throughout the paper.

Let $\mathcal{G}$ be a non-empty connected directed graph. Let $\mathcal{V} = \mathcal{V}(\mathcal{G})$ and $\mathcal{E} = \mathcal{E}(\mathcal{G})$ be the vertex and oriented edge sets respectively. For every edge e, let $i(e)$ and $t(e)$ denote the initial and the terminal vertex of e, respectively. We can define a length distance d on $\mathcal{G}$ by introducing a length function $\ell : \mathcal{E} \to \mathbb{R}$ which assigns a positive real number $\ell(e)$ to each edge $e \in \mathcal{E}$.

Example 2.1 (Infinite Graph). Consider a graph $\mathcal{G}$ formed from one vertex and a countably infinite number of edges.

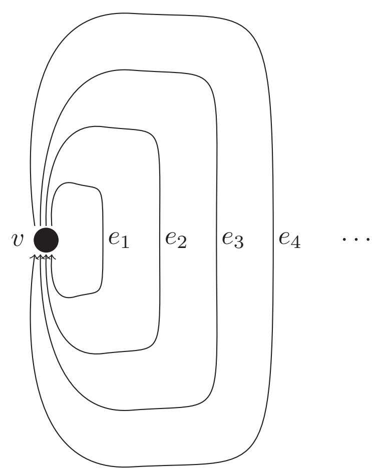


Figure 1. A single vertex $\mathcal{V} = \{v\}$ and infinitely many edges $\mathcal{E} = \{e_n\}_{n=1}^{\infty}$.

A *path* in $\mathcal{G}$ corresponds to a sequence of edges $p = e_1 \dots e_n$ for which $t(e_j) = i(e_{j+1})$, for $1 \leq j < n$ and we denote its length by $\ell(p) = \sum_{j=1}^{n} \ell(e_j)$.

Let $\mathcal{P}_{\mathcal{G}}(x, R) = \{p = e_1 \dots e_n : i(e_1) = x, \ell(p) \leq R\}$ denote the set of all such paths of length at most R starting at $x \in \mathcal{V}(\mathcal{G})$. We denote its cardinality by $N_{\mathcal{G}}(x, R) = \#\mathcal{P}_{\mathcal{G}}(x, R)$.

DEFINITION 2.2. We define the *volume entropy* of $(\mathcal{G}, \ell, x)$ as

$$h(\mathcal{G}, \ell, x) = \limsup_{R \to +\infty} \frac{1}{R} \log N_{\mathcal{G}}(x, R).$$

However, we need to make further assumptions on the length function ℓ for $h(\mathcal{G}, \ell, x)$ to be finite. To see this, consider the graph $\mathcal{G}$ in Example 2.1 which has a single vertex and an infinite number of edges, and assume that the lengths don't tend to infinity. Then for R sufficiently large, $N_{\mathcal{G}}(x, R) = \infty$ and thus $h(\mathcal{G}, \ell, x) = \infty$.

We summarise below the properties of the graph that are needed in the proof.

Graph Hypotheses. Henceforth, we shall consider graphs with finite vertex set $\mathcal{V}$ and a countable edge set $\mathcal{E}$. Furthermore we require that $\mathcal{E}$ and the associated length function satisfy the following properties:

(H1) For all $\sigma > 0$ we have $\sum_{e \in \mathcal{E}} e^{-\sigma \ell(e)} < \infty$;
(H2) For all edges $e, e' \in \mathcal{E}$ there exists a path in $\mathcal{G}$ which starts with e and ends with e'[1]; and
(H3) There does not exist a $d > 0$ such that

$$\{\ell(c) : c \text{ is a closed path}\} \subset d\mathbb{N}.$$

Under the above hypotheses, the volume entropy $h = h(\mathcal{G}, \ell, x)$ does not depend on the choice of base point x.

LEMMA 2.3. *If the graph $\mathcal{G}$ satisfies (H1) and (H2) then* $0 < h < \infty$.

PROOF. By assumption (H2), and the pigeonhole principle applied to $\mathcal{V}$, there exist a path connecting the base point x to some vertex v and two closed paths, c_1 and c_2, which pass through v. By considering all possible concatenations of these closed paths it is clear that there exists $b > 0$ such that $N_{\mathcal{G}}(x, R) \geq 2^{\lfloor R/b \rfloor}$ for all $R > 0$ and hence $h \geq \frac{\log 2}{b} > 0$.

To see that h is finite we can formally write

$$\sum_{p \in \mathcal{P}_{\mathcal{G}}(x,R)} e^{-\sigma \ell(p)} \leq \sum_{n=1}^{\infty} \left(\sum_{e \in \mathcal{E}} e^{-\sigma \ell(e)} \right)^n, \tag{2.1}$$

for $\sigma > 0$, where the Right Hand Side involves all possible sums of edge lengths. Using (H1) one can see that for $\sigma = \sigma_0$ sufficiently large $\sum_{e \in \mathcal{E}} e^{-\sigma \ell(e)} < 1$ and thus the geometric series on the Right Hand Side of (2.1) converges. In particular, since h is easily seen to be the absicssa of convergence of the series on the Left Hand Side of (2.1) we see that $h \leq \sigma_0 < +\infty$, as required. □

Our main result for $\mathcal{G}$ is the following asymptotic for the growth of paths.

THEOREM 2.1. If the graph $\mathcal{G}$ satisfies (H1),(H2) and (H3) then there exists a constant $C > 0$ such that $N_{\mathcal{G}}(x, R) \sim Ce^{hR}$, i.e.,

$$\lim_{R \to +\infty} \frac{N_{\mathcal{G}}(x, R)}{e^{hR}} = C.$$

[1] A slightly weaker assumption would be to require that for a sufficiently large finite subset $\mathcal{E}_0 \subset \mathcal{E}$, for every $e, e' \in \mathcal{E}_0$ there exists a path in $\mathcal{G}$ which starts with e and ends with e'

The proof follows the lines of the classical proof of the prime number theorem. In particular, it is based on the use of a Tauberian theorem (in Section 5). This, in turn, depends on the properties of the complex function $\eta_{\mathcal{G}}(z)$, the Laplace transform of $N_{\mathcal{G}}(x, R)$ (defined in Section 4). The function $\eta_{\mathcal{G}}(z)$ is analysed using matrices introduced in the next section. In the special case of finite graphs, the asymptotic in Theorem 2.1 could be easily deduced using ideas in [**13**] for finite matrices.

Remark 2.2. Without hypothesis (H3) this theorem may not hold. For example, even in the case of finite graphs, if we consider the graph $\mathcal{G}$ with a single vertex and two edges of length 1, then $N_{\mathcal{G}}(x, R) = 2^{\lfloor R \rfloor}$ for all $R > 0$. In this case, the limit in Theorem 2.4 does not converge.

3. Countable Matrices

In this section consider a graph $\mathcal{G}$ and length function ℓ which satisfy hypotheses (H1)-(H3). Let us order the edge set $\mathcal{E} = (e_a)_{a \in \mathbb{N}}$ by non-decreasing length and write $\ell(a) := \ell(e_a)$, $a \in \mathbb{N}$.

Definition 3.1. We can associate to $\mathcal{G}$ the infinite matrix M_0 defined by

$$M_0(a, b) = \begin{cases} 1 & \text{if } t(a) = i(b), \\ 0 & \text{otherwise.} \end{cases}$$

For each $z \in \mathbb{C}$ we define the matrix M_z by $M_z(a, b) = M_0(a, b)e^{-z\ell(b)}$ for $a, b \in \mathcal{E}$.

Let $P(n, a, b)$ denote the set of paths in $\mathcal{G}$ consisting of n edges, starting with edge e_a and ending with edge e_b. It then follows from formal matrix multiplication that for any $n \geq 1$, we can write the $(a, b)^{th}$ entry of the n^{th} power of the matrix as:

$$M_z^n(a, b) = e^{z\ell(a)} \sum_{p \in P(n+1, a, b)} e^{-z\ell(p)}, \qquad (3.1).$$

which will be finite by hypothesis (H1).

Given any matrix $L = (L(a, b))_{a,b=1}^{\infty}$ with $\sup_a \sum_b |L(a, b)| < +\infty$ we can associate to L a bounded linear operator $\widehat{L} : \ell^\infty(\mathbb{C}) \to \ell^\infty(\mathbb{C})$ by

$$\widehat{L}(\underline{u}) = \Big(\sum_{b=1}^{\infty} L(a, b)u_b \Big)_{a=1}^{\infty} \text{ where } \underline{u} = (u_b)_{b=1}^{\infty} \in \ell^\infty(\mathbb{C}).$$

In particular, by hypothesis (H1), when $Re(z) > 0$ we can associate to M_z a bounded operator $\widehat{M_z} : \ell^\infty(\mathbb{C}) \to \ell^\infty(\mathbb{C})$ by

$$\widehat{M_z}(\underline{u}) = \Big(\sum_{b=1}^{\infty} M_z(a, b)u_b \Big)_{a=1}^{\infty}.$$

To proceed, we would like to understand the domain of meromorphicity of the linear operator $(I - \widehat{M_z})^{-1} : \ell^\infty(\mathbb{C}) \to \ell^\infty(\mathbb{C})$, where I denotes the identity operator. To this end, we shall make use of an idea by Hofbauer and Keller in [**4**], where they observe that the invertibility of certain operators of the above form depends only on the determinant of an associated finite matrix.

Fix $\epsilon > 0$ and, for convenience, assume also $h > \epsilon$. Given $k \geq 1$, we can truncate the matrix M_z to the $k \times k$ matrix $A_z = (M_z(i,j))_{i,j=1}^k$. Then we can then write

$$M_z = \begin{pmatrix} A_z & B_z \\ C_z & D_z \end{pmatrix}$$

where, in particular, $D_z = (M_z(i+k, j+k))_{i,j=1}^\infty$. Again, we can interpret $I - \widehat{D_z}$ as a bounded linear operator on $\ell^\infty(\mathbb{C})$ and write $(I - \widehat{D_z})^{-1} = \sum_{m=0}^\infty \widehat{D_z}^m$ if the operator $\widehat{D_z}$ has norm $\|\widehat{D_z}\| < 1$. In particular, this is true when $Re(z) \geq \epsilon$ for k sufficiently large, since by (H1) we have

$$\|\widehat{D_z}\| \leq \sup_{n\in\mathbb{N}} \sum_{m=1}^\infty |D_z(n,m)| \leq \sum_{m=1}^\infty e^{-Re(z)\ell(m+k)} \leq \sum_{m=1}^\infty e^{-\epsilon\ell(m+k)} < 1. \qquad (3.2)$$

Writing $\ell^\infty(\mathbb{C})$ as the corresponding direct sum of two subspaces, we can then easily verify that

$$I - \widehat{M_z} = \begin{pmatrix} I - \widehat{A_z} - \widehat{B_z}(I - \widehat{D_z})^{-1}\widehat{C_z} & -\widehat{B_z}(I - \widehat{D_z})^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} I & 0 \\ -\widehat{C_z} & I - \widehat{D_z} \end{pmatrix}. \qquad (3.3)$$

Let us denote the $k \times k$ matrix $W_z := A_z + B_z(I - D_z)^{-1}C_z$, where each entry is given by a convergent series. By (3.3), whenever $\det(I - W_z) \neq 0$ then we see that $I - \widehat{M_z}$ is invertible, with inverse

$$(I - \widehat{M_z})^{-1} = \begin{pmatrix} I & 0 \\ (I - \widehat{D_z})^{-1}\widehat{C_z} & (I - \widehat{D_z})^{-1} \end{pmatrix} \begin{pmatrix} (I - \widehat{W_z})^{-1} & (I - \widehat{W_z})^{-1}\widehat{B_z}(I - \widehat{D_z})^{-1} \\ 0 & I \end{pmatrix}. \qquad (3.4)$$

This leads to the following result.

LEMMA 3.2. *The operator* $(I - \widehat{M_z})^{-1}$ *has an analytic extension to* $Re(z) > 0$ *except when* $\det(I - W_z) = 0$.

PROOF. This follows from the identity (3.4) and since the $\epsilon > 0$ used in the above construction can be chosen arbitrarily small. □

4. Complex functions

We can now introduce a complex function whose analytic properties will be useful in deriving our asymptotic estimates for $N_{\mathcal{G}}(x, R)$. Fix $x \in \mathcal{V}$.

DEFINITION 4.1. We can formally define the complex function

$$\eta_{\mathcal{G}}(z) = \int_0^\infty e^{-zR} dN_{\mathcal{G}}(x, R) = \sum_{p \in P(x)} e^{-z\ell(p)}, \quad z \in \mathbb{C},$$

where $P(x) = \{p = e_1 \cdots e_n : n \geq 0, i(e_1) = x\}$ is the set of paths in $\mathcal{G}$ starting at x.

We first observe that $\eta_{\mathcal{G}}(z)$ converges to an analytic function for $Re(z) > h$, by virtue of Definition 2.2. In order to construct a meromorphic extension of $\eta_{\mathcal{G}}(z)$ we shall relate $\eta_{\mathcal{G}}(z)$ to the matrix M_z. For $Re(z) > 0$, we define:

(a) $\underline{w}(z) = (\chi_{\mathcal{E}_x}(e_j)e^{-z\ell(j)})_{j=1}^{\infty} \in \ell^1(\mathbb{C})$ where $\chi_{\mathcal{E}_x}$ denotes the characteristic function of the set $\mathcal{E}_x = \{e \in \mathcal{E} : i(e) = x\}$ of edges whose initial vertex is x; and

(b) $\underline{1} = (1)_{j=1}^{\infty} \in \ell^{\infty}(\mathbb{C})$ is the vector all of whose entries are equal to 1,

then we can formally rewrite $\eta_{\mathcal{G}}(z)$ as

$$\begin{aligned} \eta_{\mathcal{G}}(z) = \sum_{p \in P(x)} e^{-z\ell(p)} &= \underline{w}(z) \cdot \Big(\sum_{n=0}^{\infty} \widehat{M_z}^{\,n} \Big) \underline{1} \\ &= \underline{w}(z) \cdot \Big(I - \widehat{M_z} \Big)^{-1} \underline{1}, \end{aligned} \tag{4.1}$$

where $w \cdot v = \sum_{j=1}^{\infty} w_j v_j$ for $w \in \ell^1(\mathbb{C})$ and $v \in \ell^{\infty}(\mathbb{C})$. Observe that for $Re(z) > 0$ we have $\underline{w}(z) \in \ell^1(\mathbb{C})$ by (H1). In particular, by Lemma 3.2 the expression in (4.1) extends to $Re(z) > 0$, and the locations of the poles are given by those z such that the finite rank operator $(I - W_z)$ is not invertible. Moreover, we can easily write

$$\eta_{\mathcal{G}}(z) = \frac{\phi(z)}{\det(I - W_z)} \tag{4.2}$$

where $\phi(z)$ is holomorphic on $Re(z) > 0$.

Proposition 4.1. $\eta_{\mathcal{G}}(z)$ has a meromorphic extension to $Re(z) > 0$.

Proof. Observe that $\det(I - W_z)$ is the sum of a countable number of holomorphic functions which uniformly converge on any compact domain in $Re(z) > 0$ and hence $\det(I - W_z)$ is holomorphic. The result follows from the identity (4.2). □

Let $\epsilon < h$. By (3.2) we can choose k large enough such that $(I - \widehat{D_z})$ is invertible, on the half plane $Re(z) \geq \epsilon$. Recall that a non-negative $n \times n$ matrix M is *irreducible* if for all i, j satisfying $1 \leq i, j \leq n$ there exists a natural number m such that $(M^m)_{i,j} > 0$.

Lemma 4.2. *Let $\sigma > 0$. Then W_σ is a non-negative irreducible matrix. Furthermore, W_σ has a simple maximal positive eigenvalue $\rho(\sigma) = \rho(W_\sigma)$, which depends analytically on σ and satisfies $\rho'(\sigma) < 0$.*

Proof. Recall that $W_\sigma = A_\sigma + B_\sigma(I - D_\sigma)^{-1}C_\sigma$, which by construction is a non-negative matrix. We can also deduce that the matrix W_σ is irreducible. To see this, note that by assumption (H2), for all $1 \leq i, j \leq k$, there exists some path of length n starting with edge e_i and ending with edge e_j. Such a path can be broken up into sub-paths of two types. The first type consists of those paths that stay completely within $\{e_1, \cdots, e_k\}$, and the second type which consists of those paths that initially enter the complement $\mathcal{E} - \{e_1, \cdots, e_k\}$ and finally leave at their end. Note that $W_\sigma^n(i,j)$ is a sum including powers of $A_\sigma(i,j)$ (corresponding to sub-paths of the first type) and $B_\sigma(I - D_\sigma)^{-1}C_\sigma$ (corresponding to sub-paths of the second type), where the powers are less than or equal to n. Hence $W_\sigma^n(i,j) > 0$.

We can now apply the Perron-Frobenius theorem (see [**3**]) to deduce that the maximal positive eigenvalue $\rho(\sigma) > 0$ for W_σ exists and that W_σ has associated positive left and right eigenvectors $u(\sigma)$ and $v(\sigma)$ (which we normalise so that $u(\sigma) \cdot v(\sigma) = 1$). By differentiating the eigenvalue equations for $u(\sigma)$ and $v(\sigma)$, one can show that

$$\rho'(\sigma) = u(\sigma) \cdot W'_\sigma v(\sigma) < 0,$$

where W'_σ is the matrix with entries $W'_\sigma(i,j) = \frac{dW_x(i,j)}{dx}(\sigma) < 0$ for all i,j (see [**14**] for a similar argument). □

PROPOSITION 4.2. *h is a simple pole of $\eta_{\mathcal{G}}(z)$.*

PROOF. For z in a neighbourhood of h, we denote by $\rho(z)$ the perturbed eigenvalue of W_z. We can write $\det(I - W_z) = (1-\rho(z))\Pi_{i=2}^{k}(1-\lambda_i(z))$, where the $\lambda_i(z)$ denote the other eigenvalues of W_z. Since the $\lambda_i(z)$ are bounded away from 1 for z near h (by the Perron-Frobenius theorem and standard perturbation theory), $\phi(h) \neq 0$ and $\rho'(h) \neq 0$ (by Lemma 4.2), we can conclude that $(z-h)\eta_{\mathcal{G}}(z)$ converges to a non-zero constant, as z tends to h. □

PROPOSITION 4.3. *$\eta_{\mathcal{G}}(z)$ has no poles other than h on the line $Re(z) = h$.*

PROOF. Suppose for a contraction that there exists another pole at $h + it$ ($t \neq 0$). Let c be any closed path and choose an integer $k_c > k$ such that the edges of c have index smaller than k_c. Then construct the $k_c \times k_c$ matrices W_z. From equation (4.2) we see that $\det(I - W_{h+it}) = 0$, and thus 1 is an eigenvalue for W_{h+it} and W_h. Furthermore, we can see that $\rho(W_h) = 1$ since otherwise $\eta_{\mathcal{G}}(z)$ has a pole at $c > h$, contradicting Definition 2.2.

Next observe that $|W_{h+it}(a,b)| \leq W_h(a,b)$ for all $1 \leq a,b \leq k$. Since $\rho(W_{h+it}) \geq 1 = \rho(W_h)$, we can apply Wielandt's theorem (see [**3**]) which allows us to conclude that $\rho(W_{h+it}) = \rho(W_h) = 1$ and that there exists a diagonal matrix D, whose non-zero entries have unit modulus such that $W_{h+it} = DW_hD^{-1}$, and thus for all n we have $W_{h+it}^n = DW_h^nD^{-1}$.

Suppose that the closed path c begins with some edge e_a and consists of n edges. One can check that $W_{h+it}^n(a,a) = W_h^n(a,a)$ (since $W_{h+it}^n = DW_h^nD^{-1}$) and that $e^{(h+it)\ell(c)}$ is one of the terms in the left hand sum. However, this can only hold if t is such that $\ell(c)t = 2\pi m_c$ for some non-zero integer m_c. As c was arbitrary, the above construction implies that for all closed paths c, $\ell(c) \in d\mathbb{N}$ with $d = 2\pi/t$ which contradicts (H3). □

5. Proof of Theorem 2.1

We can complete the proof using a similar approach to Parry in [**13**], where he considered only finite matrices. In particular, we will use the following formulation of the Ikehara–Wiener Tauberian theorem [**2**] applied to our counting function, $N_{\mathcal{G}}(x,R)$.

THEOREM 5.1 (Ikehara–Wiener Tauberian theorem). *Let $A : \mathbb{R}^+ \to \mathbb{R}^+$ be a monotonic, non-decreasing function and formally denote $\eta(z) := \int_0^\infty e^{-zR}dA(R)$, for $z \in \mathbb{C}$. Then suppose that $\eta(z)$ has the following properties:*

1. *there exists some $a > 0$ such that $\eta(z)$ is analytic on $Re(z) > a$;*
2. *$\eta(z)$ has a meromorphic extension to a neighbourhood of the half-plane $Re(z) \geq a$;*
3. *a is a simple pole for $\eta(z)$, i.e., $C = \lim_{z \searrow a}(z-a)\eta(z) > 0$; and*
4. *the extension of $\eta(z)$ has no poles on the line $Re(z) = a$ other than a.*

Then $A(R) \sim Ce^{aR}$ as $R \to +\infty$.

From the results in the previous section we see that the $\eta_{\mathcal{G}}(s)$ satisfies the assumptions of Theorem 5.1 with $a = h$ and so we have proved Theorem 2.1.

6. Translation surfaces

In this section we will consider a definition of volume entropy for translation surfaces and prove asymptotic results using the work developed in the previous sections.

Definition 6.1. A translation surface X is a compact surface with a flat metric except at a finite set $\Sigma = \{x_1, \dots, x_n\}$ of singular points with cone angles $2\pi(k(x_i) + 1)$, where $k(x_i) \in \mathbb{N}$, for $i = 1, \dots, n$.

A path which does not pass through singularities is a locally distance minimizing geodesic if it is a straight line segment. This includes geodesics which start and end at singularities, known as *saddle connections*. We will consider oriented saddle connections.

Geodesics can change direction if they go through a singular point, and a pair of line segments ending and beginning, respectively, at a singular point form a geodesic if the angle between them is at least π. Thus a locally distance minimising geodesic (of length R) on a translation surface X with singularity set Σ, is a curve $\gamma : [0, R] \to X$ satisfying the following conditions:

- There exist $0 \le t_1 < ... < t_n \le R$, where $n \ge 0$, such that $\gamma(t_i) \in \Sigma$;
- For $t_i < t < t_{i+1}$ $\gamma(t) \in X \backslash \Sigma$;
- $\gamma : (t_i, t_{i+1}) \to X \backslash \Sigma$ is a geodesic segment (possibly a saddle connection);
- The smallest angle between $\gamma|_{(t_{i-1}, t_i)}$ and $\gamma|_{(t_i, t_{i+1})}$ is at least π (cf. [**1**], Lemma 2.1).

Let $\mathcal{S} = \{s_1, s_2, ...\}$ be the set of oriented saddle connections ordered by non-decreasing lengths.

Definition 6.2. We define a saddle connection path $p = (s_{i_1}, ..., s_{i_n})$ to be a finite string of oriented saddle collections $s_{i_1}, ..., s_{i_n}$ which form a geodesic path.

We denote by $\ell(p) = \ell(s_1) + \ell(s_2) + \cdots + \ell(s_n)$ the sum of the lengths of the constituent saddle connections. We let $i(p), t(p) \in \Sigma$ denote the initial and terminal singularities, respectively, of the saddle connection path p.

Example 6.3 (Square tiled surfaces [**16**]). We can consider the square-tiled surfaces by identifying opposite sides of arrangements of a finite number of copies of the unit square (Figure 2). The values of the lengths of the saddle connections are of the form $\{\sqrt{n^2 + m^2} : (n, m) \in \mathbb{Z}^2 - (0, 0) \text{ co-prime}\}$.

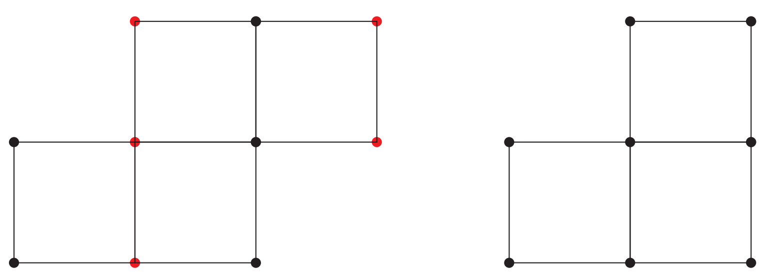


Figure 2. (i) A square tiled surface formed from four tiles; (ii) a square tiled surface formed from three tiles.

We now turn our attention to defining a notion of volume entropy for translation surfaces in terms of the growth of the volume of a ball as its radius tends to infinity. By analogy with the definition of volume entropy for Riemannian manifolds (Definition 1.1) we can consider the rate of growth of balls in the universal cover $\widetilde{X}$ of X.

DEFINITION 6.4. Let $\tilde{x} \in \widetilde{X}$ and consider a ball $B(\tilde{x}, R) \subset \widetilde{X}$ of radius $R > 0$ with centre $\widetilde{x}$. We define the volume entropy of X to be

$$h = h(X) := \limsup_{R \to +\infty} \frac{1}{R} \log \mathrm{Vol}_{\widetilde{X}}(B(\tilde{x}, R))$$

where $\mathrm{Vol}_{\widetilde{X}}$ denotes the natural volume on $\widetilde{X}$.

Definition 6.4 is closely related to the definition of Dankwart [**1**], which was formulated in terms of orbital counting. As in the case of the definitions of volume entropy for Riemannian manifolds and finite metric graphs, h is independent of the choice of $\tilde{x}$. For convenience, we can take $\tilde{x}$ to be the lift of a singularity $x \in \Sigma$.

It is also possible to interpret this definition in terms of X rather than $\widetilde{X}$. To this end we have the following definition.

DEFINITION 6.5. Let $m_R(y)$ be number of distinct geodesic arcs in X from x to y of length at most R.

We can now rewrite $\mathrm{Vol}_{\widetilde{X}}(B(\tilde{x}, R)) = \int_X m_R(y) d\mathrm{Vol}_X(y)$ (see Figure 3). For economy of notation we will write $V(x, R) := \mathrm{Vol}_{\widetilde{X}}(B(\tilde{x}, R))$.

Let $x \in \Sigma$ be a singularity, then we define

$$\pi(x, R) := \{p : i(p) = x \text{ and } l(p) \leq R\}$$

to be the number of saddle connection paths starting at x of length less than or equal to R.

LEMMA 6.6. *Let X be a translation surface and fix a singularity $x \in \Sigma$ and let $2\pi(k(x) + 1)$ be the cone angle of x. Then for $R > 0$,*

$$V(x, R) = (k(x) + 1)\pi R^2 + \sum_{p \in \pi(x,R)} k(t(p))\pi(R - \ell(p))^2,$$

where the singularity at the end of path p has cone angle $2\pi(k(t(p)) + 1)$.

PROOF. The volume contributed by the geodesics starting from x which do not pass through a singularity is given by $(k(x)+1)\pi R^2$, where $2\pi(k(x)+1)$ is the cone angle at x. On the other hand, the contribution to the volume by those geodesics γ which pass through one or more singularities comes when the geodesic leaves its last singularity at time $\ell(p) < R$, say. It can exit in one of $2\pi k(p)$ directions. Then the total volume of such γ is given by $k(t(p))\pi(R - (\ell(p))^2$. □

We shall now prove asymptotic results for translation surfaces using the analysis developed for infinite graphs.

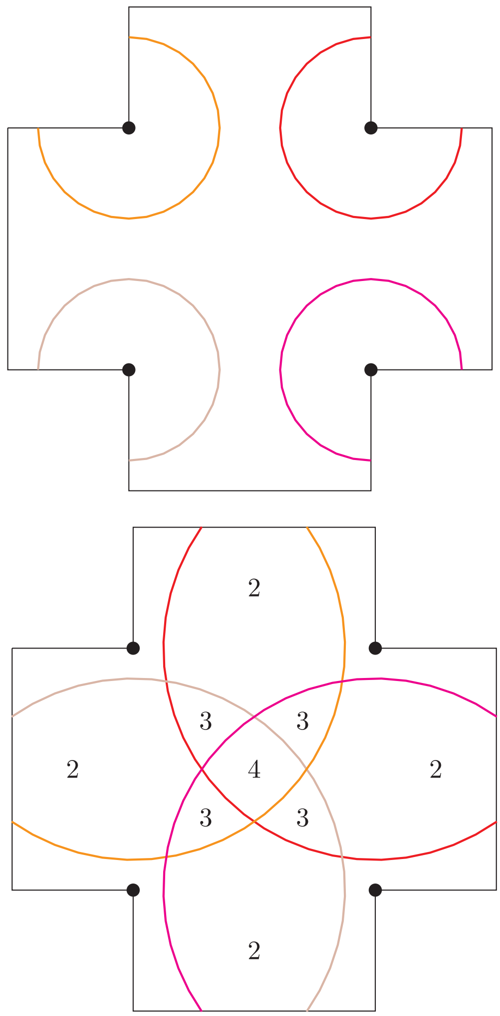


FIGURE 3. (i) A small ball centred at a singularity; (ii) As the radius R increases the ball overlaps with itself (and the values of the multiplicity function $m_R(\cdot)$ are indicated).

DEFINITION 6.7. We can associate to X the countable matrix M_0, indexed by $\mathcal{S}$, defined by

$$M_0(s, s') = \begin{cases} 1 & \text{if ss}' \text{ form a saddle connection path,} \\ 0 & \text{otherwise.} \end{cases}$$

For each $z \in \mathbb{C}$ we define the matrix M_z by $M_z(s, s') = M_0(s, s')e^{-z\ell(s')}$ for $s, s' \in \mathcal{S}$.

In order that the matrices have the same properties that served us well for graphs, we require specific features of a translation surface.

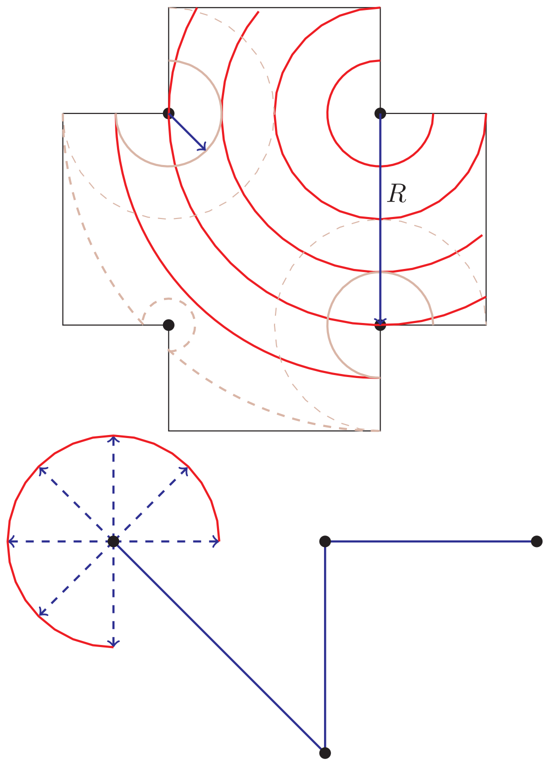


FIGURE 4. (i) The radii of the projections $B(\tilde{x}, R)$ of balls in the universal cover are concatenations of saddle connections followed by a radial line segment from a singularity; (ii) A heuristic figure illustrating that the boundary of $B(\tilde{x}, R)$ will consist of the union of circular arcs centred on singularities reached via concatenations of saddle connections

Translation Hypotheses. Henceforth, we shall consider translation surfaces whose countable set of saddle connections is denoted by $\mathcal{S}$. Moreover, we require that $\mathcal{S}$ and the lengths of the saddle connections satisfy the following properties:

(T1) For all $\sigma > 0$ we have $\sum_{s\in\mathcal{S}} e^{-\sigma\ell(s)} < +\infty$;
(T2) For any directed saddle connections $s, s' \in \mathcal{S}$ there exists a saddle connection path beginning with s and ending with s'; and
(T3) There does not exist a $d > 0$ such that

$$\{\ell(c) : c \text{ is a closed saddle connection path}\} \subset d\mathbb{N}.$$

We claim that the above hypotheses hold for all translation surfaces.

Property (T1) follows from the lower bound in following result (see [**10**] and [**11**]).

PROPOSITION 6.1. Let X be a translation surface and let $N(X, L)$ denote the number of saddle connections on X of length less than or equal to L. Then there exists constants $0 < c_1 < c_2 < \infty$ such that

$$c_1 L^2 \le N(X, L) \le c_2 L^2,$$

for L sufficiently large.

To see that Hypotheses (T2) and (T3) hold for all translation surfaces, we require the following result in [**1**] which we restate for our purposes here.

PROPOSITION 6.2. Let X be a translation surface. If $s, s' \in \mathcal{S}$ are oriented saddle connections then there exists a saddle connection path which starts with s and ends with s'.

Hypothesis (T2) follows immediately from this fact.

To show Hypothesis (T3) holds for all surfaces we first note that if the lengths of all closed geodesics were an integer multiple of some constant d, then the length of every saddle connection would be an integer multiple of $d/2$. To see this, let s be any saddle connection on X. If $i(s) = t(s)$ then s is a closed geodesic and so we are done. If $i(s) \neq t(s)$ then by Proposition 6.2, there exists a closed saddle connection path c_i such that c_i passes through $i(s)$ and that $\bar{s}c_i s$ forms a saddle connection path (where $\bar{s}$ is the saddle connection s with reversed orientation). Similarly, there exists a closed saddle connection path c_t which starts and ends at $t(s)$, such that $sc_t\bar{s}$ forms a saddle connection path. Note that the concatenation $sc_t\bar{s}c_i$ is also a closed saddle connection path of length $2\ell(s) + \ell(c_t) + \ell(c_i)$ and so by Hypothesis (T3), $\ell(s) \in (d/2)\mathbb{N}$. Let us now assume for a contradiction that (T3) does not hold and, in particular, the above property holds for the saddle connection lengths.

Using results in [**12**], X contains an embedded cylinder C (the product of a circle with an interval) whose boundaries consist of a single saddle connection or multiple parallel saddle connections. We now aim to construct a countable family of triangles using this cylinder (Figure 5). Fix two singularities x and y, one from

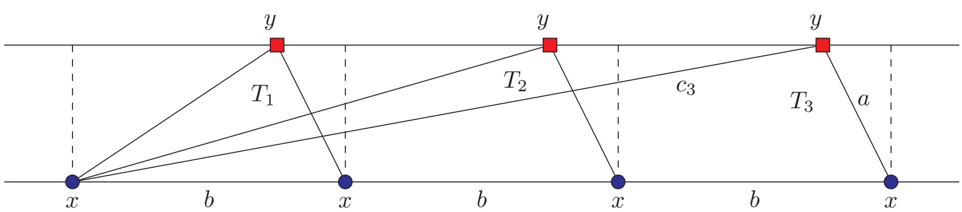


FIGURE 5. Three copies of a cylinder on X with two singularities on separate boundaries represented by circles and squares. The corresponding triangles T_1, T_2 and T_3 are also drawn.

each boundary. Let b denote the union of saddle connections which form the boundary of the cylinder connecting x to itself. Let a be one of the saddle connection connecting x to y across the cylinder such that the angle between a and b is acute. Then consider the unique saddle connection c_n connecting x to y which is defined to be the third side in a triangle T_n whose other edges are b concatenated with itself n times and a. By hypothesis each edge has length which is an integer multiple of $d/2$. However, by elementary Euclidean geometry we can show that this cannot hold for all sufficiently large n, giving the required contradiction.

To derive an asymptotic estimate for $V(x,R)$ we can associate the complex function

$$\eta_X(z) = \int_0^\infty e^{-zR} dV(x,R).$$

Let $\mathcal{P}(x) := \{p : i(p) = x\}$ denote the set of saddle connection paths starting at x. We can rewrite $\eta_X(z)$ as follows:

$$\begin{aligned}
\eta_X(z) &= \int_0^\infty e^{-zR}\left(\frac{d}{dR}\mathrm{Vol}(B(x,R))\right) dR \\
&= 2\pi(k(x)+1)\int_0^\infty e^{-zR} R dR + 2\pi \sum_{p\in\mathcal{P}(x)} k(t(p)) \int_{\ell(p)}^\infty e^{-zR}(R-\ell(p))dR \\
&= 2\pi(k(x)+1)\int_0^\infty e^{-zR} R dR + 2\pi \sum_{p\in\mathcal{P}(x)} k(t(p)) e^{-z\ell(p)} \int_0^\infty e^{-zR} R dR \\
&= \frac{2\pi}{z^2}(k(x)+1) + \frac{2\pi}{z^2}\sum_{p\in\mathcal{P}(x)} k(t(p)) e^{-z\ell(p)} \\
&= \frac{2\pi}{z^2}(k(x)+1) + \frac{2\pi}{z^2}\underline{v}(z)\cdot(I-\widehat{M_z})^{-1}\underline{u},
\end{aligned} \tag{6.1}$$

where $\underline{u} = (k(t(s_j)))_{j=1}^\infty \in \ell^\infty(\mathbb{C})$ and $\underline{v}(z) = (\chi_{\mathcal{E}_x}(s_j)e^{-z\ell(s_j)})_{j=1}^\infty \in \ell^1(\mathbb{C})$, where $\chi_{\mathcal{E}_x}$ denotes the characteristic function of the set $\mathcal{E}_x = \{s \in \mathcal{S} : i(s) = x\}$ of saddle connections starting from the singularity $x \in \Sigma$.

LEMMA 6.8. *The function $\eta_X(z)$ is analytic for $Re(z) > h$ and has a meromorphic extension to $Re(z) > 0$. Moreover $\eta_X(z)$ has a simple pole at $z = h$ and no other poles on $Re(z) = h$.*

PROOF. We can apply the analysis of $(I-\widehat{M_z})^{-1}$ in Section 4 to (6.1), where we use hypotheses (T1)-(T3) in place of (H1)-(H3). □

We can now apply Theorem 5.1 to deduce the following.

THEOREM 6.3. There exists a $C > 0$ such that $V(x,R) \sim Ce^{hR}$ as $R \to +\infty$, i.e.,

$$\lim_{R\to+\infty} \frac{V(x,R)}{e^{hR}} = C.$$

Typically $C = C(x)$ will depend on the choice of x.

There is a closely related result for counting the number of geodesic arcs $N_X(x,y,R)$ starting at $x \in \Sigma$ and finishing at $y \in \Sigma$.

PROPOSITION 6.4. There exists a $D > 0$ such that $N_X(x,y,R) \sim De^{hR}$ as $R \to +\infty$, i.e.,

$$\lim_{R\to+\infty} \frac{N_X(x,y,R)}{e^{hR}} = D.$$

PROOF. The proof simply requires replacing the function $\eta_X(z)$ by the function

$$\eta_N(z) = \int_0^\infty e^{-zR} dN_X(x,y,R) = \underline{v}(z)\cdot(I-\widehat{M_z})^{-1}\underline{w},$$

where $\underline{w} = (\chi_{\mathcal{F}}(s_i))_{i=1}^\infty$, with $\chi_{\mathcal{F}}(s)$ denoting the characteristic function for the set $\mathcal{F} = \{s \in \mathcal{S} : t(s_i) = y\}$ of saddle connections ending at the singularity y and $u(z)$

was defined after equation (6.1). Again the properties of $(I - \widehat{M_z})^{-1}$ allow one to apply Theorem 5.1 to deduce the result. $\square$

REMARK 6.5. We conclude with some final remarks.

(1) It is not necessary for the ball in Theorem 6.3 to be centered at a singularity. Let $y \in X - \Sigma$ and let G be the set of geodesics g, from y to a singularity, such that g has length $\ell(g)$. Order G by non-decreasing lengths. We define a matrix P where
  (a) the rows are indexed by such geodesics g and the columns are indexed by the oriented saddle connections s;
  (b) the non-zero entries correspond to pairs g, s such that:
    (i) The singularity $t(g)$ at the end of g is the same as that $i(s)$ at the start of the saddle connection s; and
    (ii) The geodesic g and saddle connection s have an angle of at least π between them.
  (c) The non-zero entries are $P_z(g,s) = e^{-z\ell(s)}$.

One can then modify the complex function to $\eta_X(z) = \frac{2\pi}{z^2}(k(x)+1) + \frac{2\pi}{z^2}\underline{v}_p(z) \cdot \widehat{P_z}(I - \widehat{M_z})^{-1}\underline{u}$, where $\underline{v}_p(z) = (e^{-z\ell(g)})_{g \in G} \in \ell^1(\mathbb{C})$ and then continue the proof as in Theorem 6.3.

(2) Theorem 6.3 also follows as a corollary of Theorem 6.4 by using a simple approximation argument. In particular, this shows that $C{=}D \int_0^\infty e^{-u}u^2 du$.

(3) Let $L(x,R)$ be the total circumference of a circle centred at x and whose radius is a geodesic of length R. The same approach as in the proof of Theorem 6.3 (or an approximation argument as in item 2) would give an asymptotic formula of the form: There exists $E > 0$ such that $L(x,R) \sim Ee^{hR}$, as $R \to +\infty$.

(4) Eskin and Rafi have announced a closely related asymptotic result to Theorem 6.3 for closed geodesics on X. By studying zeta functions $\zeta_X(z)$ instead of eta functions $\eta_X(z)$ they show that the number of closed geodesics of length at most $R > 0$ is asymptotic to e^{hR}/hR as $R \to +\infty$.

References

[1] K. Dankwart, On the large scale geometry of flat surfaces, Dissertation, Bonn, 2010.

[2] W. Ellison and F. Ellison, *Prime numbers*, A Wiley-Interscience Publication, John Wiley & Sons, Inc., New York; Hermann, Paris, 1985. MR814687

[3] F. R. Gantmacher, *The theory of matrices. Vols. 1, 2*, Translated by K. A. Hirsch, Chelsea Publishing Co., New York, 1959. MR0107649

[4] F. Hofbauer and G. Keller, *Zeta-functions and transfer-operators for piecewise linear transformations*, J. Reine Angew. Math. **352** (1984), 100–113, DOI 10.1515/crll.1984.352.100. MR758696

[5] P. Hubert and T. A. Schmidt, *An introduction to Veech surfaces*, Handbook of dynamical systems. Vol. 1B, Elsevier B. V., Amsterdam, 2006, pp. 501–526, DOI 10.1016/S1874-575X(06)80031-7. MR2186246

[6] S. Lim, *Minimal volume entropy for graphs*, Trans. Amer. Math. Soc. **360** (2008), no. 10, 5089–5100, DOI 10.1090/S0002-9947-08-04227-X. MR2415065

[7] A. Manning, *Topological entropy for geodesic flows*, Ann. of Math. (2) **110** (1979), no. 3, 567–573, DOI 10.2307/1971239. MR554385

[8] G. A. Margulis, *Certain applications of ergodic theory to the investigation of manifolds of negative curvature* (Russian), Funkcional. Anal. i Priložen. **3** (1969), no. 4, 89–90. MR0257933

[9] G. A. Margulis, *On some aspects of the theory of Anosov systems*, Springer Monographs in Mathematics, Springer-Verlag, Berlin, 2004. With a survey by Richard Sharp: Periodic orbits of hyperbolic flows; Translated from the Russian by Valentina Vladimirovna Szulikowska. MR2035655

[10] H. Masur, *Lower bounds for the number of saddle connections and closed trajectories of a quadratic differential*, Holomorphic functions and moduli, Vol. I (Berkeley, CA, 1986), Math. Sci. Res. Inst. Publ., vol. 10, Springer, New York, 1988, pp. 215–228, DOI 10.1007/978-1-4613-9602-4_20. MR955824

[11] H. Masur, *The growth rate of trajectories of a quadratic differential*, Ergodic Theory Dynam. Systems **10** (1990), no. 1, 151–176, DOI 10.1017/S0143385700005459. MR1053805

[12] H. Masur and J. Smillie, *Hausdorff dimension of sets of nonergodic measured foliations*, Ann. of Math. (2) **134** (1991), no. 3, 455–543, DOI 10.2307/2944356. MR1135877

[13] W. Parry, *An analogue of the prime number theorem for closed orbits of shifts of finite type and their suspensions*, Israel J. Math. **45** (1983), no. 1, 41–52, DOI 10.1007/BF02760669. MR710244

[14] W. Parry and S. Tuncel, *Classification problems in ergodic theory*, London Mathematical Society Lecture Note Series, vol. 67, Cambridge University Press, Cambridge-New York, 1982. Statistics: Textbooks and Monographs, 41. MR666871

[15] M. Pollicott and M. Urbanski, Asymptotic counting in conformal dynamical systems, Memoirs Amer. Math. Soc., to appear, arXiv 1704.06896 (2017).

[16] G. Schmithüsen, *Examples for Veech groups of origamis*, The geometry of Riemann surfaces and abelian varieties, Contemp. Math., vol. 397, Amer. Math. Soc., Providence, RI, 2006, pp. 193–206, DOI 10.1090/conm/397/07473. MR2218009

[17] A. Zorich, *Flat surfaces*, Frontiers in number theory, physics, and geometry. I, Springer, Berlin, 2006, pp. 437–583, DOI 10.1007/978-3-540-31347-2_13. MR2261104

Department of Mathematics, Warwick University, Coventry, CV4 7AL, UK
Email address: p.m.g.colognese@warwick.ac.uk

Department of Mathematics, Warwick University, Coventry, CV4 7AL, UK
Email address: masdbl@warwick.ac.uk

Contemporary Mathematics
Volume **744**, 2020
https://doi.org/10.1090/conm/744/14982

Dynamically affine maps in positive characteristic

Jakub Byszewski, Gunther Cornelissen, and Marc Houben
with Appendix B by the authors and Lois van der Meijden

dedicated to the memory of Sergiy Kolyada

Abstract. We study fixed points of iterates of dynamically affine maps (a generalisation of Lattès maps) over algebraically closed fields of positive characteristic p. We present and study certain hypotheses that imply a dichotomy for the Artin–Mazur zeta function of the dynamical system: it is either rational or non-holonomic, depending on specific characteristics of the map. We also study the algebraicity of the so-called tame zeta function, the generating function for periodic points of order coprime to p. We then verify these hypotheses for dynamically affine maps on the projective line, generalising previous work of Bridy, and, in arbitrary dimension, for maps on Kummer varieties arising from multiplication by integers on abelian varieties.

1. Introduction

We consider so-called *dynamically affine maps*, a concept in algebraic dynamics introduced by Silverman [**43**, §6.8] in order to unify various interesting examples, such as Chebyshev and Lattès maps, cousins of which occur in complex dynamics under the name of "finite quotients of affine maps" or "rational maps with flat orbifold metric" [**35**]. We will only consider the case of a ground field of positive characteristic $p > 0$. (Most of our methods would simplify considerably in characteristic zero and lead to results of a rather different flavour.) Before we present the definition, we will illustrate by approximative pictures (constructed in Mathematica [**51**], using the function `RandomInteger` for randomisation) what distinguishes the dynamics of such maps from that of other polynomials maps and random maps.

1.1. A compilation of (restrictions of) maps. Let $f\colon S \to S$ denote a map from a finite set S to itself. It can be represented by a directed graph D_f (sometimes called the "function digraph" of f), with vertices labelled by elements of S and an arrow from a vertex x to a vertex y occurring precisely if $f(x) = y$. In Figure 1, we plotted the graphs corresponding to two random such maps where S is a set with $7^3 + 1$ elements.

2010 *Mathematics Subject Classification.* 37P55, 14K15, 37C30.

Key words and phrases. Fixed points, Artin–Mazur zeta function, dynamically affine map, recurrence sequence, holonomic sequence, natural boundary.

JB gratefully acknowledges the support of National Science Center, Poland under grant no. 2016/23/D/ST1/01124.

Now consider a rational function $f\colon \mathbf{P}^1(\overline{\mathbf{F}}_p) \to \mathbf{P}^1(\overline{\mathbf{F}}_p)$ defined over $\mathbf{F}_p$ (in this subsection we assume for convenience that $p \neq 2$). To represent f pictorially, consider the restrictions $f|_{\mathbf{F}_{p^N}}: \mathbf{P}^1(\mathbf{F}_{p^N}) \to \mathbf{P}^1(\mathbf{F}_{p^N})$ for various N. In Figure 2, we plotted the graph of the polynomial function $x \mapsto x^2+1$ for various p and N, and in Figure 3, we did the same for $x \mapsto x^2-2$. At first sight, the graph for a random map looks similar to the graph for $x \mapsto x^2+1$, but the graph for $x \mapsto x^2-2$ looks much more structured. This is no coincidence; Figure 3 represents the graph of restrictions of a dynamically affine map, whereas Figure 2 does not.

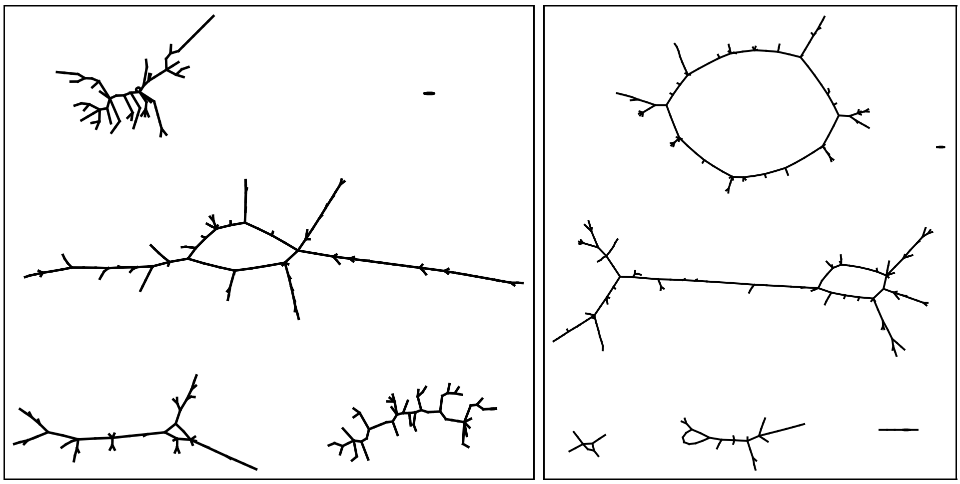


FIGURE 1. Graph of two random maps on a set with 7^3+1 elements

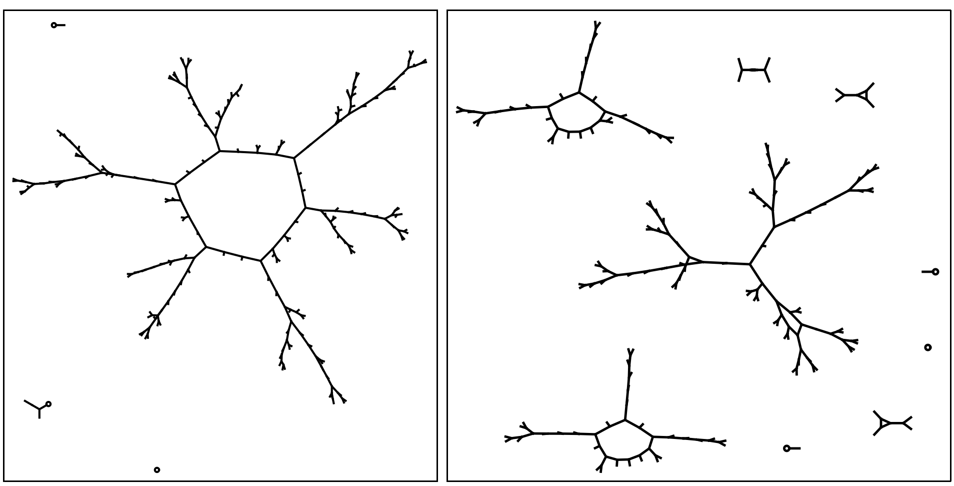


FIGURE 2. Graphs of $x \mapsto x^2+1$ on $\mathbf{P}^1$ over a field with 7^3 and 17^2 elements (left to right)

A common feature of all function digraphs is that their connected components are cycles (consisting of periodic points) with attached trees. What is different in Figure 3 is the symmetry in the attached trees; this is well-understood for the polynomial $x \mapsto x^2 - 2$, which relates to the Lucas–Lehmer test and failure of the Pollard rho method of factorisation, see, e.g. [**38**, **50**]. Let us mention one further result [**29**, Thm. 1.5 & Example 7.2]: for the graph of a quadratic polynomial with integer coefficients, the value of

$$\liminf_{p \to +\infty} \#\{x \in \mathbf{F}_p \text{ belongs to a cycle of } D_{f \bmod p}\}/p$$

is 0 for $x^2 + 1$ but 1/4 for $x^2 - 2$.

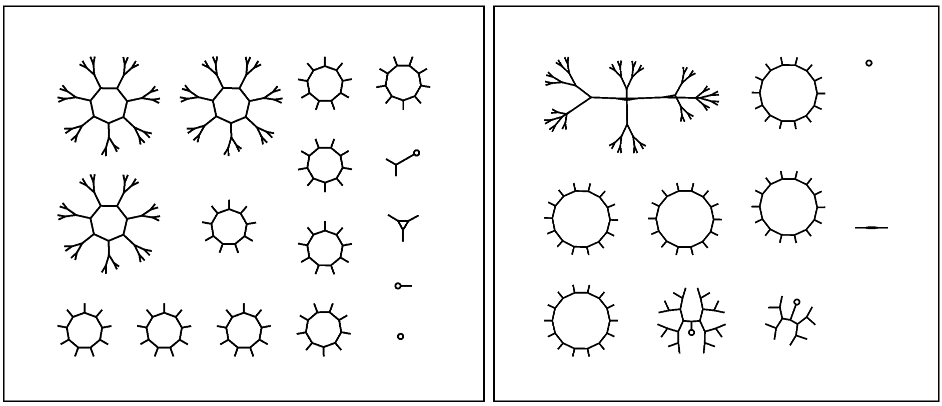


FIGURE 3. Graphs of $x \mapsto x^2 - 2$ on $\mathbf{P}^1$ over a field with 7^3 and 17^2 elements (left to right)

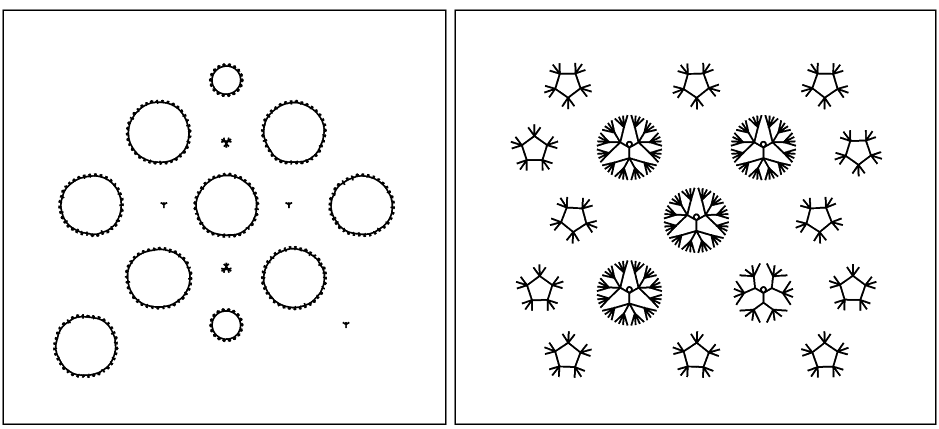


FIGURE 4. Graphs of the Lattès map arising from doubling modulo inversion on the elliptic curve $E\colon y^2 = x(x-1)(x-2)$ over a field with 11^3 and 23^2 elements (left to right)

To explain what is special about the dynamically affine map $x \mapsto x^2 - 2$ as opposed to the polynomial map $x \mapsto x^2 + 1$, notice that $x^2 - 2 = T_2(x)$, where T_2 is the normalised Chebyshev polynomial of the second kind, defined by

$T_n(x+x^{-1}) = x^n+x^{-n}$. This reveals a hidden group structure: the map arises from the group endomorphism $\sigma\colon \mathbf{G}_m \to \mathbf{G}_m, x \mapsto x^2$ on the multiplicative group $\mathbf{G}_m$ after quotienting on both sides by the automorphism group $\Gamma = \{1, \gamma\}$ generated by the inversion $\gamma\colon \mathbf{G}_m \to \mathbf{G}_m, z \mapsto z^{-1}$ that commutes with σ. That $x^2 + 1$ (for $p \neq 2, 3$) is not special in this sense follows from the classification of dynamically affine maps on $\mathbf{P}^1$ [**9**].

We perform a similar construction using another algebraic group, the elliptic curve $E\colon y^2 = x(x-1)(x-2)$, and the doubling map $\sigma\colon E \to E, P \mapsto 2P$. After taking the quotient by $\Gamma = \langle P \mapsto -P\rangle$, we find a so-called *Lattès map* $\mathbf{P}^1 \to \mathbf{P}^1$ which we have graphed over various finite fields in Figure 4. Again, we see a very structured picture, rather different from Figure 1 and Figure 2.

We will not dwell any longer on the study of iterations of maps on finite sets, both random and "polynomial over finite fields"—a rich subject in itself—but rather switch to our main object of study: dynamically affine maps over algebraically closed fields of positive characteristic.

1.2. What is a dynamically affine map? Let V be an algebraic variety over an algebraically closed field K of characteristic p and $f\colon V \to V$ a morphism. We make the following assumption throughout:

- **(C)** The map f is *confined*, i.e. the number of fixed points f_n of the n-th iterate f^n of f is finite for all n.

DEFINITION. A morphism $f\colon V \to V$ of an algebraic variety V over K is called *dynamically affine* if there exist:

- (i) a connected commutative algebraic group $(G, +)$;
- (ii) an *affine morphism* $\psi\colon G \to G$, that is, a map of the form
$$g \mapsto \psi(g) = \sigma(g) + h,$$
where $\sigma \in \mathrm{End}(G)$ is a confined isogeny (i.e. a surjective homomorphism with finite kernel) and $h \in G(K)$;
- (iii) a finite subgroup $\Gamma \subseteq \mathrm{Aut}(G)$; and
- (iv) a morphism $\iota\colon \Gamma\backslash G \to V$ that identifies $\Gamma\backslash G$ with a Zariski-dense open subset of V

such that the following diagram commutes:

$$\begin{array}{ccc} G & \xrightarrow{\psi} & G \\ \downarrow \pi & & \downarrow \pi \\ \Gamma\backslash G & \longrightarrow & \Gamma\backslash G \\ \downarrow \iota & & \downarrow \iota \\ V & \xrightarrow{f} & V. \end{array} \tag{1}$$

REMARK. In this paper, we adhere to the convention that a dynamically affine map consists of all the given (fixed) data in the definition, so that we can refer to the constituents $(G, \psi, \sigma, h, \Gamma, \iota)$ directly. The same map f might arise from different sets of data, and in our sense be a different dynamically affine map despite being the same map on V.

EXAMPLE. As explained above, the map $\mathbf{P}^1 \to \mathbf{P}^1, x \mapsto x^2 - 2$ is dynamically affine for the data $(G = \mathbf{G}_m, \sigma\colon x \mapsto x^2, h = 1, \Gamma = \langle z \mapsto z^{-1}\rangle, \iota\colon \Gamma\backslash\mathbf{G}_m \simeq \mathbf{A}^1 \hookrightarrow \mathbf{P}^1)$

(written multiplicatively); its restrictions (to certain finite fields) were represented in Figure 3.

The map $\mathbf{P}^1 \to \mathbf{P}^1, x \mapsto (x^4-4x^2-4)/(4x(x-1)(x-2))$ is dynamically affine for the data $(G=E, \sigma\colon P \mapsto 2P, h=0_E, \Gamma = \langle P \mapsto -P\rangle, \iota\colon \Gamma\backslash E \cong \mathbf{P}^1 \to \mathbf{P}^1)$, where E is the elliptic curve $y^2 = x(x-1)(x-2)$; its restrictions were represented in Figure 4.

Remark. We have slightly modified Silverman's definition [**43**, §6.8] of a dynamically affine map. Instead of assuming confinedness of σ, Silverman imposes the condition $\deg(\sigma) \geqslant 2$ (as in Erëmenko's classification theorem [**20**]). As long as G is one-dimensional and $K = \overline{\mathbf{F}}_p$, the definitions are equivalent.

In a general setup one could assume merely that σ is an isogeny and only require f to be confined. This reduces, after some case distinctions, to the case where σ is a confined isogeny, so we choose to put the latter property in the definition.

1.3. Counting fixed points, orbits, and the dynamical Artin–Mazur zeta function. A natural way to begin a quantitative analysis of a discrete dynamical system such as iteration of a map $f\colon V \to V$ is to consider the sequence (f_n) given by the number of fixed points of the n-th iterate of f. Confinedness implies that this is a well-defined sequence of integers, and we can form the (full) Artin–Mazur dynamical zeta function ([**2**], [**45**, §4]) defined as

$$\zeta_f(z) := \exp\left(\sum_{n\geqslant 1} f_n \frac{z^n}{n}\right). \tag{2}$$

We consider this a priori as a formal power series, but the question of convergence in a neighbourhood of $z=0$ (equivalent to f_n growing at most exponentially in n) is interesting, and we study this in Appendix A.

Counting fixed points and closed orbits is related: if P_ℓ denotes the number of closed orbits of length ℓ, then $f_n = \sum_{\ell|n} \ell P_\ell$, and there is an "Euler product"

$$\zeta_f(z) = \prod_C \frac{1}{1-z^{\ell(C)}}, \tag{3}$$

where the product runs over the closed orbits C.

It is interesting to understand the nature of the function $\zeta_f(z)$ (Smale [**45**, Problem 4.5]); Artin and Mazur [**2**, Question 2 on p. 84]). For example, rationality or algebraicity of $\zeta_f(z)$ means that there is an easy recipe to compute all f_n from a finite amount of data (in the rational case, it implies that (f_n) is linearly recurrent). Zeta functions of more general dynamical systems can:

– *be rational*: e.g. for "Axiom A" diffeomorphisms by Manning [**32**, Cor. 2], for rational functions of degree $\geqslant 2$ on the Riemann sphere by Hinkkanen [**27**, Thm. 1], for the Weil zeta function (when f is the Frobenius map on a variety defined over a finite field) by Dwork [**19**] and Grothendieck [**26**, Cor. 5.2], for endomorphisms of real tori [**4**, Thm. 1], and when f_n replaced by the Lefschetz number of f^n [**45**];
– *be algebraic but not rational*: e.g. when f is an orientation preserving surface homeomorphism and f_n is replaced by the Nielsen number of f^n by Pilyugina and Fel'shtyn [**36**], [**21**, Thm. 36];

- *be transcendental*: e.g. for restrictions of shifts by Bowen and Lanford [**7**, §3–4] and for separable dynamically affine maps on $\mathbf{P}^1(\overline{\mathbf{F}}_p)$ by Bridy [**8**, Thm. 1], [**9**, Thm. 1.2 & 1.3];
- *have an essential singularity*: e.g. for some flows by Gallavotti [**23**, §4];
- *have a natural boundary*: e.g. for certain beta-transformations by Flatto, Lagarias and Poonen [**22**, Thm. 2.4], for some $\mathbf{Z}^d$-actions ($d \geqslant 2$) by Lind [**31**], for some flows by Pollicott [**37**, §4] and Ruelle [**40**], for a "random" such zeta function by Buzzi [**10**], for some explicit automorphisms of solenoids by Bell, Miles, and Ward [**6**], and for most endomorphisms of abelian varieties in characteristic $p > 0$ by the first two authors [**11**].

Following the philosophy of [**11**], we will also study "tame dynamics" via the so-called *tame zeta function* defined by

$$\zeta_f^*(z) := \exp\left(\sum_{p \nmid n} f_n \frac{z^n}{n}\right), \tag{4}$$

summing only over n that are not divisible by p. Tame and "full" dynamics are related by the formulae in (5) below, but the tame zeta function tends to be better behaved. In Appendix B, we give some explicit expressions for the tame zeta function of several dynamically affine maps on $\mathbf{P}^1$.

1.4. Main results. Bridy studied the zeta function for dynamically affine maps on $V = \mathbf{P}^1$. The main results in [**9**, Thm. 1.2 & 1.3] imply that if f is dynamically affine for $V = \mathbf{P}^1$ and $K = \overline{\mathbf{F}}_p$, then $\zeta_f(z)$ is transcendental over $\mathbf{C}(z)$ (the field of rational functions with complex coefficients) if and only if f is separable; otherwise $\zeta_f(z)$ is rational. Bridy's full result applies to all K; the proof uses a case-by-case analysis (see Table 1 in Appendix B below) and is based on the relation between transcendence and automata theory. This starkly contrasts with the fact that in characteristic zero all dynamically affine maps have a rational zeta function (a much more general result by Hinkkanen was quoted above).

In this paper, we prove a strengthening of Bridy's result. For this, we need some further concepts. Let $f \colon V \to V$ be a dynamically affine map.

Definition. An endomorphism $\sigma \in \mathrm{End}(G)$ is said to be *coseparable* if $\sigma^n - 1$ is a separable isogeny for all $n \in \mathbf{Z}_{>0}$. A dynamically affine map f is called coseparable if the associated isogeny σ is coseparable.

Remark. In [**11**], we called a coseparable endomorphism of an abelian variety "very inseparable" and showed that this implies inseparability [**11**, 6.5(ii)]. However, it is not true that coseparable dynamically affine maps are inseparable in general. For example, if f is the map $f \colon \mathbf{P}^1 \to \mathbf{P}^1, x \mapsto tx$ for $t \in K$ transcendental over $\overline{\mathbf{F}}_p$, then f is both coseparable and separable (a more general statement is given in [**9**, Thm. 1.3]).

Definition. A holomorphic function on a connected open subset $\Omega \subseteq \mathbf{C}$ is said to have a *natural boundary* along $\partial\Omega$ if it has no holomorphic continuation to any larger such $\Omega' \supsetneq \Omega$ [**41**, §6]. We call a function $F(z)$ *root-rational* if $F(z)^t \in \mathbf{C}(z)$ for some $t \in \mathbf{Z}_{>0}$. We call $F(z)$ *holonomic* if it satisfies a nontrivial linear differential equation with coefficients in $\mathbf{C}(z)$.

Since algebraic functions are holonomic [**49**, Thm. 6.4.6], the following is indeed a strengthening of Bridy's result. At the same time, it shows that "tame" dynamics is better behaved.

THEOREM A. *Assume* $f\colon \mathbf{P}^1 \to \mathbf{P}^1$ *is a dynamically affine map.*

(i) *If f is coseparable, $\zeta_f(z)$ is a rational function; otherwise, $\zeta_f(z)$ is not holonomic; more precisely, it is a product of a root-rational function and a function admitting a natural boundary along its circle of convergence.*
(ii) *For all f, $\zeta_f^*(z)$ is root-rational; equivalently, it is algebraic and satisfies a first order differential equation over $\mathbf{C}(z)$.*

We mention an amusing corollary of Theorem A: although $\zeta_f(z)$ is in general not holonomic, the *pair* $(\zeta_f(z), \zeta_{f^p}(z))$ always satisfies a simple differential equation; see Corollary 2.4 for a precise statement.

Rather than using results from automata theory, we prove Theorem A essentially relying on a method of Mahler (see [**5**]). We structure the proof abstractly, showing the result for dynamically affine maps (in any dimension) that satisfy certain hypotheses **(H1)**–**(H4)** (see Section 3), and then verify these for $V = \mathbf{P}^1$.

We give a more general discussion of when the hypotheses hold or fail, in this way producing the first higher-dimensional examples of dynamically affine maps in positive characteristic with nontrivial Γ where we understand the nature of the dynamical zeta function. Recall that the quotient of an abelian variety A by the group $\Gamma = \{[\pm 1]\}$ is called a *Kummer variety.*

THEOREM B. *Let V denote a Kummer variety arising from an abelian variety A, and let $f\colon V \to V$ denote the dynamically affine map induced by the multiplication-by-m map $\sigma = [m]$ for some integer $m \geqslant 2$. Then $\zeta_f^*(z)$ is root-rational. The function $\zeta_f(z)$ is not holonomic if m is coprime to p and rational otherwise.*

REMARK. We use the word "Kummer variety" for the variety $V = \Gamma\backslash A$ that, for $\dim A > 1$, is singular at points in the finite subset $\Gamma\backslash A[2]$ of V, but the name is sometimes used for the minimal resolution $\tilde{V}$ of V. Since the set of singular points is finite and stable by f, the map f can be seen as a birational map $f\colon \tilde{V} \dashrightarrow \tilde{V}$ with locus of indeterminacy stable by f, and the above theorem can be interpreted as a statement about the periodic points of this birational map outside the preimage of the singular points.

REMARK. The non-holonomicity shows that the sequence (f_n) of number of fixed points of the iterates of f is somewhat "complex", but it does not mean that f_n is "uncomputable". As a matter of fact, the results in [**12**] say that for f an endomorphism of an algebraic group there exists a formula expressing f_n in terms of a linear recurrent sequence and two specific periodic sequences of integers that control a p-adic deviation of f_n from being linearly recurrent. These data can in principle be computed by breaking up the algebraic group into abelian varieties, tori, vector groups, and semisimple groups. Similarly, one can in principle trace through our proofs to compute f_n for dynamically affine maps satisfying our hypotheses.

We finish the introduction by mentioning a few possibilities for future research.

– The relation between fixed points and closed orbits may be used to study the distribution of closed orbit lengths (analogously to the prime number theorem).

Because of the analytic nature of the function $\zeta_f(z)$ revealed by our results, one cannot in general use standard Tauberian methods. We have studied this question via a different route for maps on abelian varieties [**11**] and for maps on general algebraic groups [**12**] (which covers the case of dynamically affine maps with trivial Γ, h, and ι, but is more general, since we do not require the group G to be commutative). It would be interesting to extend this to general dynamically affine maps.

- We have no good understanding of the dynamical zeta function of general rational functions on $\mathbf{P}^1$ that are not dynamically affine, e.g. $x \mapsto x^2+1$ in characteristic $p \geqslant 5$ (see [**8**, Question 2]). It would be interesting to investigate the nature of the (tame) zeta function for such examples.
- Inhowfar the hypotheses **(H1)**–**(H4)** are necessary to reach the conclusion of the main theorem merits attention, since they are extracted from a "method of proof" rather than intrinsic.
- In general, V may be singular. It is interesting to study whether V admits a resolution to which f extends as a morphism, and the relation between the zeta function of that extended morphism and the zeta function of f. This is nontrivial already for Kummer surfaces (where, for $p > 2$, the minimal resolution is a K3 surface, and hence has trivial étale fundamental group [**28**, pp. 3–6]).

The structure of the paper is as follows: After some generalities, we introduce the hypotheses in Section 3 and prove the main result, conditional on the hypotheses, in the following section. Then, in Section 5 we discuss the validity of the hypotheses in various settings (giving examples and counterexamples). The main theorems then follow immediately from these results. In the first appendix, we consider the radius of convergence of $\zeta_f(z)$, and in the second appendix, we compute a collection of examples of tame zeta functions of dynamically affine maps.

2. Generalities

Relations between zeta functions.

PROPOSITION 2.1. *The tame and full dynamical zeta function are related by the following equalities of formal power series:*

$$\zeta_f^*(z) = \frac{\zeta_f(z)}{\sqrt[p]{\zeta_{f^p}(z^p)}}, \qquad \zeta_f(z) = \prod_{i\geqslant 0} \sqrt[p^i]{\zeta^*_{f^{p^i}}(z^{p^i})}. \tag{5}$$

PROOF. For the first equality, note that

$$\log \zeta_f^*(z) = \sum_{n\geqslant 1} f_n \frac{z^n}{n} - \frac{1}{p}\sum_{m\geqslant 1} f_{pm}\frac{z^{pm}}{m} = \log\left(\zeta_f(z)\zeta_{f^p}(z^p)^{-1/p}\right).$$

The second equality follows by applying the first one to the functions f^{p^i} for $i \in \mathbf{Z}_{\geqslant 0}$. ☐

REMARK 2.2. A useful computational fact is the following: if $f\colon S \to S$ is a map and S decomposes as a union $S = S_1 \cup S_2$ with $f(S_1) \subseteq S_1$ and $f(S_2) \subseteq S_2$, then

$$\zeta_f(z) = \frac{\zeta_{f|_{S_1}}(z)\cdot \zeta_{f|_{S_2}}(z)}{\zeta_{f|_{S_1\cap S_2}}(z)},$$

and similarly for $\zeta_f^*(z)$.

Recurrences. We recall some well-known facts (see e.g. [**11**, §1]). If $(a_n)_{n\geqslant 1}$ is a sequence of complex numbers, then the ordinary generating function $\sum_{n\geqslant 1} a_n z^n$ is rational if and only if the sequence is linear recurrent, and if and only if there exist $\lambda_i \in \mathbf{C}^\times$ and polynomials $p_i \in \mathbf{C}[z]$ such that

$$a_n = \sum_{i=1}^{r} p_i(n)\lambda_i^n \tag{6}$$

for sufficiently large n. The statement that the zeta function

$$F(z) = \exp\left(\sum_{n\geqslant 1} a_n \frac{z^n}{n}\right) \tag{7}$$

is rational is stronger: this happens if and only if Equation (6) holds for all $n \in \mathbf{Z}_{>0}$ with the $p_i(n)$ replaced by integers m_i independent of n. The λ_i occurring in (6) are called the *roots* of the recurrence, the polynomials p_i their *multiplicities*. We say that (a_n) satisfies the *dominant root assumption* if there is a unique root λ_i of maximal absolute value, possibly with multiplicity $\neq 1$.

For a zeta function $F(z)$ in (7), we may consider its tame variant

$$F^*(z) = \exp\left(\sum_{\substack{n\geqslant 1\\ p\nmid n}} a_n \frac{z^n}{n}\right).$$

It follows from the formula

$$F^*(z) = F(z) \cdot \left(\prod_{j=0}^{p-1} F(e^{\frac{2i\pi j}{p}} z)\right)^{-1/p} \tag{8}$$

that if $F(z)$ is rational, then $F^*(z)$ is root-rational.

Algebraicity properties and differential equations. If a formal power series $F(z)$ satisfies a nontrivial linear differential equation over $\mathbf{C}(z)$, it is said to be *holonomic*. If $F(z)$ is algebraic over $\mathbf{C}(z)$, it is holonomic [**49**, Thm. 6.4.6]. On the other hand, if $F(z)$ converges on some nontrivial open disc D and has natural boundary along ∂D, then it cannot be holonomic, since a holonomic function has only finitely many singularities (for a precise statement, see [**48**, 4(a)]).

The equivalence statement in Theorem A(ii) is implied by the following lemma, which is certainly well-known, but for which we were unable to find a convenient reference. (A more general result can be found in [**49**, Exercise 6.62] together with an argument attributed to B. Dwork and M. F. Singer.)

LEMMA 2.3. *An algebraic function $F(z) \in \mathbf{C}(\!(z)\!)$ is root-rational if and only if f satisfies a first order homogeneous differential equation $F'(z) = R(z)F(z)$ with $R(z) \in \mathbf{C}(z)$.*

PROOF. First assume that $F(z)$ is root-rational, i.e. $F(z) = q(z)^k$ with $q(z) \in \mathbf{C}(z)$, $k \in \mathbf{Q}$. We may assume that $q(z) \neq 0$, and then $F(z)$ satisfies the equation $F(z)' = R(z)F(z)$ with $R(z) = \frac{kq'(z)}{q(z)}$.

The converse direction can be proven by direct integration and partial fraction expansion of $R(z)$, but we give a somewhat different argument. Assume that $F(z)$

satisfies the equation $F(z)' = R(z)F(z)$ with $R(z) \in \mathbf{C}(z)$, where we may assume $R(z) \neq 0$. Let $P \in \mathbf{C}(z)[t]$ be the minimal polynomial of $F(z)$ over $\mathbf{C}(z)$. Write $P = t^d + a_{d-1}(z)t^{d-1} + \cdots + a_0(z)$ with $a_i(z) \in \mathbf{C}(z)$. Differentiating the equation $P(F(z)) = 0$ gives

$$P^D(F(z)) + P'(F(z))F(z)' = 0, \tag{9}$$

where $P^D = \sum_{i=0}^d a_i'(z)t^i$ is obtained from P by differentiating the coefficients and $P' = \sum_{i=0}^d ia_i(z)t^{i-1}$ is the usual derivative of P. Substituting $F'(z) = R(z)F(z)$ into (9), we see that F is a root of the polynomial $P^D + tR(z)P'$, which is a polynomial of degree d with leading coefficient $dR(z)$, and hence

$$P^D + tR(z)P' = dR(z)P.$$

Comparing the coefficients at t^i for $i = 0, \ldots, d-1$, we see that each $a_i(z)$ satisfies the equation

$$a_i'(z) = (d-i)R(z)a_i(z),$$

which differs from the equation satisfied by $F(z)$ only by a multiplicative constant. Comparing these solutions gives $a_i(z) = c_iF(z)^{d-i}$ for some $c_i \in \mathbf{C}$. If $a_i(z) = 0$ for all $i \in \{1, \ldots, d\}$, we get $F(z) = 0$. Otherwise, for some i we have $a_i(z) \neq 0$, and $F(z) = (c_i^{-1}a_i(z))^{1/(d-i)}$ is root-rational. □

Thus, Theorem A(ii) immediately implies the result alluded to in the introduction:

Corollary 2.4. *If $f\colon \mathbf{P}^1 \to \mathbf{P}^1$ is a dynamically affine map, then the* pair *of zeta functions $(F_1(z), F_2(z)) = (\zeta_f(z), \zeta_{f^p}(z))$ satisfies a nonlinear first order differential equation*

$$F_1'(z)F_2(z^p) - F_1(z)F_2'(z^p)z^{p-1} = R(z)F_1(z)F_2(z^p)$$

for some rational function $R(z) \in \mathbf{C}(z)$, regardless of whether or not f is coseparable.

Proof. The root-rationality of $\zeta_f^*(z)$ implies that it satisfies a differential equation of the form $(\zeta_f^*(z))' = R(z)\zeta_f^*(z)$ for some rational function $R(z) \in \mathbf{C}(z)$. The result follows by taking derivatives in the first identity in (5). □

3. Introduction of the general hypotheses

Let $f\colon V \to V$ be a dynamically affine map with data as in diagram (1). Denote by $\mathrm{Orb}_f(x) := \{f^n(x) \mid n \in \mathbf{Z}_{\geqslant 0}\}$ the forward orbit of $x \in V(K)$ under f. For an isogeny $\tau \in \mathrm{End}(G)$, we denote by $\deg(\tau)$ and $\deg_{\mathrm{i}}(\tau)$ the degree and inseparable degree of the field extension $K(G)/\tau_*K(G)$, respectively. Then we have

$$\#\ker(\tau) = \deg(\tau)/\deg_{\mathrm{i}}(\tau). \tag{10}$$

The following lemma, taken from [**9**, Lemma 2.4] (cf. Remark 4.2), will be crucial to control the sequence (f_n), as it allows us to express f_n in terms of kernels of isogenies on the algebraic group G. The proof will be given in Section 4.

Lemma 3.1. *Let $f\colon V \to V$ be a dynamically affine map. Consider the set*

$$C := \{x \in V(K) \mid \mathrm{Orb}_f(x) \cap \iota((\Gamma\backslash G)(K)) = \emptyset\}.$$

Then

$$f_n = (f|_C)_n + \frac{1}{|\Gamma|}\sum_{\gamma\in\Gamma} \#\ker(\sigma^n - \gamma). \tag{11}$$

Combining Lemma 3.1 with (10), we see that in order to understand the sequence (f_n) it suffices to control, for every $\gamma \in \Gamma$,

(a) the sequence $(f|_C)_n$;
(b) the "inseparable degree sequence" $\deg_i(\sigma^n - \gamma)$;
(c) the "degree sequence" $\deg(\sigma^n - \gamma)$.

Notice that the translation parameter $h \in G(K)$ no longer occurs in (11).

We now introduce the four hypotheses that we require in order to prove the main theorems. The first three hypotheses **(H1)**, **(H2)** and **(H3)** are employed to control the sequences (a), (b) and (c), respectively, while **(H4)** is a technical hypothesis that we require to avoid an unexpected cancellation of singularities in our proof of the existence of a natural boundary.

We use the following

> CONVENTION: If a hypothesis is assumed in an environment (definition, lemma, theorem, hypothesis, ...), we label the environment by this hypothesis in square brackets.

HYPOTHESIS **(H1)**. *The zeta function corresponding to $f|_C$ is rational.*

For the second hypothesis, we recall the following notion: a *discrete valuation* on a (not necessarily commutative) ring R is a map $v\colon R \to \mathbf{Z}\cup\{\infty\}$ such that for all $\tau, \tau_1, \tau_2 \in R$ we have $v(\tau) = \infty$ if and only if $\tau = 0$, $v(\tau_1\tau_2) = v(\tau_1) + v(\tau_2)$, and $v(\tau_1 + \tau_2) \geqslant \min\{v(\tau_1), v(\tau_2)\}$. It follows from these properties that $v(\tau_1 + \tau_2) = \min\{v(\tau_1), v(\tau_2)\}$ whenever $v(\tau_1) \neq v(\tau_2)$.

HYPOTHESIS **(H2)**. *Both σ and Γ belong to a subring $\mathscr{R}$ of* $\mathrm{End}(G)$ *all of whose nonzero elements are isogenies, and such that there exists a discrete valuation $v\colon \mathscr{R} \to \mathbf{Z}\cup\{\infty\}$ satisfying* $\deg_i(\tau) = p^{v(\tau)}$ *for all isogenies $\tau \in \mathscr{R}$.*

Note that the valuation v considered in **(H2)** takes only nonnegative values.

Before introducing the last two hypotheses, we set up some notation.

NOTATION 3.2. Let v be as in **(H2)**. For $m \in \mathbf{Z}_{\geqslant 0}$, we let

$$\Gamma_m := \{\gamma \in \Gamma \mid v(\gamma - 1) \geqslant m\}.$$

This defines a descending filtration of normal subgroups of Γ

$$\Gamma = \Gamma_0 \supseteq \Gamma_1 \supseteq \cdots \supseteq \Gamma_N = 1,$$

where

$$N := \max\{v(\gamma - 1) \mid \gamma \in \Gamma, \gamma \neq 1\} + 1.$$

For $m \in \mathbf{Z}_{\geqslant 0}$ we define $s_m \in \mathbf{Z}_{>0}$ to be the smallest integer such that $v(\sigma^{s_m} - \gamma_m) \geqslant m$ for some $\gamma_m \in \Gamma$; in general, s_m might not exist, but s_0 certainly does, and we will show in Lemma 4.11 that for $m > 0$ either none of the s_m exist or all do depending on whether or not f is coseparable. Write $s := s_N$ and $\tilde{\gamma} := \gamma_N$.

Hypothesis **(H3)**. [**(H2)**] *Let $m \in \mathbf{Z}_{\geqslant 0}$. If s_m exists, then*

$$\exp\left(\frac{1}{|\Gamma_m|}\sum_{\substack{n\geqslant 1\\ \gamma\in\Gamma_m}} \deg(\sigma^{s_m n} - \gamma\gamma_m^n)\frac{z^n}{n}\right) \in \mathbf{C}(z).$$

Remark 3.3. The statement of Hypothesis **(H3)** a priori depends on the choice of the elements γ_m. However, it will follow from Lemma 4.12(ii) below that it is independent of such a choice.

Hypothesis **(H4)**. [**(H2)**] *The number s exists and the sequence*

$$(\deg(\sigma^{sn} - \tilde{\gamma}^n))_{n\geqslant 1} \tag{12}$$

is a linear recurrent sequence satisfying the dominant root assumption.

Remark 3.4. If s exists and **(H3)** holds, then the sequence (12) is automatically linear recurrent. Moreover, by Lemma 4.11, s exists if and only if f is not coseparable, and the element $\tilde{\gamma} \in \Gamma$ is then unique.

We then have the following results:

Theorem 3.5. *Assume $f\colon V \to V$ is a dynamically affine map satisfying the hypotheses* **(H1)**–**(H4)**. *Then $\zeta_f(z)$ is not holonomic. More precisely, it is a product of a root-rational function and a function admitting a natural boundary along its circle of convergence.*

Theorem 3.6. *Assume $f\colon V \to V$ is a dynamically affine map satisfying the hypotheses* **(H1)**–**(H3)**. *Then $\zeta_f^*(z)$ is root-rational.*

The proofs of these theorems will be given in the next section.

4. Proofs of Theorems 3.5 and 3.6

Preliminary results on the action of Γ.

Lemma 4.1. *Let $f\colon V \to V$ be a dynamically affine map.*

(i) *There exists a group automorphism $\alpha\colon \Gamma \to \Gamma$ such that for any $\gamma \in \Gamma$, $\psi\gamma = \alpha(\gamma)\psi$ and $\sigma\gamma = \alpha(\gamma)\sigma$.*
(ii) *The map $\sigma^n - \gamma$ is an isogeny for all $n \in \mathbf{Z}_{>0}$ and $\gamma \in \Gamma$.*
(iii) *$\#(\psi^n - \gamma)^{-1}(0) = \#(\sigma^n - \gamma)^{-1}(0)$ for all $n \in \mathbf{Z}_{>0}$ and $\gamma \in \Gamma$.*

Proof. (i) That α exists as a map of sets follows from [**43**, Prop. 6.77(a)(b)]. Recall that, by assumption, σ is surjective and has finite kernel. Now, for all $\gamma_1, \gamma_2 \in \Gamma$ we have

$$\alpha(\gamma_1\gamma_2)\sigma = \sigma(\gamma_1\gamma_2) = (\sigma\gamma_1)\gamma_2 = \alpha(\gamma_1)\alpha(\gamma_2)\sigma,$$

which implies that α is a group homomorphism. For $\gamma \in \ker(\alpha)$, we have $\sigma(\gamma - 1) = 0$, and so $\operatorname{im}(\gamma - 1) \subseteq \ker(\sigma)$. Since $\ker(\sigma)$ is finite and G is connected, we must have $\operatorname{im}(\gamma - 1) = \{0\}$, and so $\gamma = 1$. This shows that α is injective, and hence bijective.

(ii) Let $\gamma \in \Gamma$ and $n \in \mathbf{Z}_{>0}$. We will show that $\sigma^n - \gamma$ has finite kernel. Suppose that $x \in G(K)$ is such that $\sigma^n(x) = \gamma(x)$. Put $\beta := \alpha^n$. Then

$$\sigma^{dn}(x) = \left(\beta^{d-1}(\gamma)\cdots\beta(\gamma)\gamma\right)(x). \tag{13}$$

Since β is injective and Γ is finite, there exists $d \in \mathbf{Z}_{>0}$ for which

$$\beta^{d-1}(\gamma) \cdots \beta(\gamma)\gamma = 1,$$

so that $(\sigma^n - \gamma)^{-1}(0) \subseteq (\sigma^{dn} - 1)^{-1}(0)$. Since by assumption σ is confined, we have that $(\sigma^{dn} - 1)^{-1}(0)$ is finite, and the desired result follows.

(iii) For every n, there exists $h_n \in G(K)$ such that $\psi^n(g) = \sigma^n(g) + h_n$ for all $g \in G(K)$. We then have

$$\#(\psi^n - \gamma)^{-1}(0) = \#(\sigma^n - \gamma)^{-1}(-h_n) = \#(\sigma^n - \gamma)^{-1}(0),$$

where in the last equality we use the fact that $\sigma^n - \gamma$ is an isogeny. □

Proof of Lemma 3.1. The proof of [**9**, Lemma 2.4] shows that

$$f_n = (f|_C)_n + \frac{1}{|\Gamma|} \sum_{\gamma \in \Gamma} \#(\psi^n - \gamma)^{-1}(0).$$

The desired result now follows from Lemma 4.1(iii). □

Remark 4.2. The claim in [**9**, Lemma 2.4] that (11) holds for dynamically affine maps using Silverman's definition (under the additional assumption that ψ is surjective), is incorrect. For example, when $V = G = E \times E$ for an elliptic curve E, $\Gamma = \{1\}$, $\sigma = [1] \times [2]$, and $h = (P, 0)$ with $P \in E(K)$ a non-torsion point, then $f_n = \psi_n = 0$, but $\ker(\sigma^n - 1) \supseteq E(K) \times \{0\}$ is infinite for all $n \in \mathbf{Z}_{>0}$. The mistake in the proof is that under the assumptions in Silverman's definition, Lemma 4.1(iii) does not need to hold (for this one needs part (ii) of the lemma, which is equivalent to σ being confined). Nevertheless, in [**9**] the result is only applied for $\dim V = 1$, where Silverman's definition implies confinedness of σ, hence none of the other results are affected.

Preliminary results on valuations.

Proposition 4.3. *Let R denote a (not necessarily commutative) ring with a discrete valuation v. Then the following statements hold for all $x, y \in R$ and $n \in \mathbf{Z}_{>0}$:*

(i) *R has no nontrivial zero divisors.*
(ii) *The characteristic of R is either zero or prime.*
(iii) *If $v(x) \geqslant 0$ and $v(y) \geqslant 0$, then $v(xy - yx) \geqslant v(x - y)$.*
(iv) *If $v(x) \geqslant 0$ and $v(y) \geqslant 0$, then $v(x^n - y^n) \geqslant v(x - y)$.*
(v) *Assume that x and y commute, $v(x) = v(y) = 0$, and $v(x - y) > 0$. Then:*
  (a) *if $\mathrm{char}(R) = 0$ and $v(\mathbf{Z} - \{0\}) = 0$, then $v(x^n - y^n) = v(x - y)$;*
  (b) *if $\mathrm{char}(R) = 0$ and $v(p) > 0$ for some prime p, then if $v(x - y) > v(p)/(p - 1)$, we have $v(x^n - y^n) = v(x - y) + v(n)$;*
  (c) *if $\mathrm{char}(R) = p > 0$, then $v(x^n - y^n) = v(x - y) \cdot |n|_p^{-1}$.*
(vi) *In cases* (b) *and* (c) *above, if $z \in R$ satisfies $v(z - 1) > 0$, then $v(z^n - 1)$ is unbounded as n ranges over $\mathbf{Z}_{>0}$.*

Proof. (i) Follows directly from the fact that the valuation $v(x)$ of x is infinite if and only if $x = 0$.

(ii) Follows from (i).

(iii) Follows from the formula $xy - yx = (x - y)x - x(x - y)$.

(iv) Let R' be the subring of R generated by x and y. Then the restriction of v to R' is a valuation on R' taking only nonnegative values. We have $x^n - y^n =$

$(y+(x-y))^n - y^n = y^n - y^n + z$, where z lies in the two-sided ideal of R' generated by $(x-y)$, and hence $v(x^n - y^n) = v(z) \geqslant v(x-y)$.

(v) Since x and y commute, we have

$$x^n - y^n = n(x-y)y^{n-1} + \sum_{k=2}^{n} \binom{n}{k} (x-y)^k y^{n-k}. \tag{14}$$

If $v(n) = 0$, then the first term has strictly smaller valuation than the second one, and hence $v(x^n - y^n) = v(x-y)$, proving case (a), as well as cases (b) and (c) for $p \nmid n$. It now suffices to consider (b) and (c) for $n = p$; the general result will then follow by induction on $v_p(n)$. For (b), the assumption on $v(x-y)$ implies that

$$v\left(\binom{p}{k}\right) + (k-1)v(x-y) > v(p) \tag{15}$$

for all $2 \leqslant k \leqslant p$. This shows that again in (14) the first term has strictly smaller valuation than the second one, which yields $v(x^p - y^p) = v(x-y) + v(p)$. For (c), note that $v(x^p - y^p) = v((x-y)^p) = pv(x-y)$.

(vi) Follows from the formula (14) with $x = z$, $y = 1$, and n an arbitrarily large power of p. □

Remark 4.4. [**(H2)**] If R as above is the endomorphism ring of a connected commutative algebraic group G over K and v is a valuation on $\mathrm{End}(G)$ satisfying [**(H2)**], then Proposition 4.3(ii) can be made slightly more explicit: the characteristic of R will then be either zero or equal to $p = \mathrm{char}(K)$. In fact, if $\ell \in \mathbf{Z}_{>0}$ is a prime and $v(\ell) > 0$, then the multiplication-by-ℓ map is either zero or an inseparable isogeny, and hence its differential, which on the tangent space at 0 is given by multiplication by ℓ, is not an isomorphism. Since the tangent space at 0 is a K-vector space, we must have $p = \mathrm{char}(K) > 0$ and $\ell = p$. This also implies that the prime p found in (v)(b) is equal to $\mathrm{char}(K)$.

Remark 4.5. The assumption that x and y commute is necessary in Proposition 4.3(v)(b). Consider the quaternion algebra $\mathbf{H}$ generated over $\mathbf{Q}$ by i, j with $i^2 = j^2 = -1$ and $ij = -ji$, and let $\mathscr{O} = \mathbf{Z} + \mathbf{Z}\, i + \mathbf{Z}\, j + \mathbf{Z}\, \frac{1+i+j+k}{2}$ be the ring of Hurwitz quaternions, which is a maximal order in $\mathbf{H}$. Consider the valuation v on $\mathscr{O}$ corresponding to the prime element $1 + i \in \mathscr{O}$. Put $x = i + 4j$ and $y = i$. Then $v(x^2 - y^2) = v(-16) = 8$, but $v(x-y) + v(2) = 6$. The assumption that x and y commute is missing from [**9**, Lemma 6.2], but the result is only applied for $y = 1$, and so other results in that reference are not affected.

Recall that $f \colon V \to V$ is a dynamically affine map with associated data as in diagram (1). Assume that f satisfies **(H2)**. In order to obtain more information about f, we will apply Proposition 4.3 to the ring $R = \mathscr{R}$ and the valuation v supplied by **(H2)**.

Lemma 4.6. [**(H2)**] *If σ is coseparable, then $\sigma^n - \gamma$ is a separable isogeny for all $n \in \mathbf{Z}_{>0}$ and $\gamma \in \Gamma$.*

Proof. Let $\gamma \in \Gamma$ and $n \in \mathbf{Z}_{>0}$. By Lemma 4.1(ii), $\sigma^n - \gamma$ is an isogeny, so it remains to show that it is separable. Applying Proposition 4.3(iv), we see that

$$v(\sigma^n - \gamma) \leqslant v(\sigma^{|\Gamma| n} - \gamma^{|\Gamma|}) = v(\sigma^{|\Gamma| n} - 1) = 0,$$

where in the last equality we use that σ is coseparable. □

PROPOSITION 4.7. [**(H1)**–**(H3)**] *If f is coseparable, then $\zeta_f(z)$ is rational and $\zeta^*_f(z)$ is root-rational.*

PROOF. If f is coseparable, then by Lemma 4.6, $\#\ker(\sigma^n - \gamma) = \deg(\sigma^n - \gamma)$ for all n and γ. The desired result for $\zeta_f(z)$ then follows by applying Lemma 3.1 together with **(H1)** and **(H3)** with $m = 0$. By the general rationality conditions in Section 2, this implies that $\zeta_{f^p}(z)$ is rational as well, and hence the result for $\zeta^*_f(z)$ follows from (5). □

REMARK 4.8. Proposition 4.7 is false if we drop the assumption of the hypotheses. In fact, if **(H2)** and **(H3)** do not hold, then $\zeta_f(z)$ may even have a natural boundary along its circle of convergence (see Example 5.9 below).

LEMMA 4.9. [**(H2)**] *If f is not coseparable, then $v(\sigma) = 0$.*

PROOF. If $v(\sigma) > 0$, then $v(\sigma^n - 1) = 0$ for all n, contradicting the assumption that σ is not coseparable. □

LEMMA 4.10. [**(H2)**] *Suppose that $n \in \mathbf{Z}_{>0}$ and $\gamma \in \Gamma$ are such that $v(\sigma^n - \gamma) \geqslant N$. Then σ^n and γ commute.*

PROOF. Since $N > 0$ and $v(\gamma) = 0$, we have $v(\sigma) = 0$. Let $\alpha \in \mathrm{Aut}(\Gamma)$ as in Lemma 4.1(i), so that $\sigma\gamma = \alpha(\gamma)\sigma$. It follows that

$$N \leqslant v(\sigma^n - \gamma) \leqslant v(\sigma^n\gamma - \gamma\sigma^n) = v((\alpha^n(\gamma) - \gamma)\sigma^n) = v(\alpha^n(\gamma) - \gamma).$$

We conclude that $\alpha^n(\gamma) = \gamma$, and hence $\sigma^n\gamma = \alpha^n(\gamma)\sigma^n = \gamma\sigma^n$. □

We will now prove the announced result on the existence of the numbers s_m defined in Notation 3.2.

LEMMA 4.11. [**(H2)**]

(i) *If f is coseparable, then none of the numbers s_m exist for $m > 0$.*
(ii) *If f is not coseparable, then all of the numbers s_m exist.*

PROOF. (i) If f is coseparable, then by Lemma 4.6 all the maps $\sigma^n - \gamma$ for $n \in \mathbf{Z}_{>0}$ and $\gamma \in \Gamma$ are separable isogenies, and hence $v(\sigma^n - \gamma) = 0$. Hence s_m do not exist for $m > 0$.

(ii) Since f is not coseparable, there is some $n \in \mathbf{Z}_{>0}$ such that $v(\sigma^n - 1) > 0$, and hence by Proposition 4.3.(vi) the values of $v(\sigma^n - 1)$ can be arbitrarily large. This proves the existence of s_m for all m. □

LEMMA 4.12. [**(H2)**] *Suppose that f is not coseparable. Let $m \in \mathbf{Z}_{\geqslant 0}$. Then:*

(i) *The set*

$$S_m := \{n \in \mathbf{Z}_{>0} \mid v(\sigma^n - \gamma) \geqslant m \text{ for some } \gamma \in \Gamma\}$$

is equal to $s_m\,\mathbf{Z}_{>0}$.

(ii) *For every $n \in \mathbf{Z}_{>0}$, $\{\gamma \in \Gamma \mid v(\sigma^{s_m n} - \gamma) \geqslant m\} = \Gamma_m\gamma_m^n$.*

PROOF. (i) By Proposition 4.3(iv), we have $v(\sigma^{s_m n} - \gamma_m^n) \geqslant v(\sigma^{s_m} - \gamma_m) \geqslant m$, so $S_m \supseteq s_m\,\mathbf{Z}_{>0}$. Now suppose to the contrary that there exists an $n \in S_m - s_m\,\mathbf{Z}_{>0}$. Then there exists a $\gamma \in \Gamma$ such that $v(\sigma^n - \gamma) \geqslant m$, and we can write $n = ds_m + r$ for $0 < r < s_m$. We obtain

$$m \leqslant v(\sigma^n - \gamma) = v(\sigma^r(\sigma^{ds_m} - \gamma_m^d) + (\sigma^r - \gamma\gamma_m^{-d})\gamma_m^d).$$

This leads to a contradiction since $v(\sigma^r(\sigma^{ds_m} - \gamma_m^d)) \geqslant m$ and $v((\sigma^r - \gamma\gamma_m^{-d})\gamma_m^d) < m$.

(ii) We know that $v(\sigma^{s_m n} - \gamma_m^n) \geqslant m$, so for any $\gamma \in \Gamma$ we have the equivalence

$$\begin{aligned} v(\sigma^{s_m n} - \gamma) \geqslant m &\iff v(\gamma - \gamma_m^n) \geqslant m \\ &\iff v(\gamma\gamma_m^{-n} - 1) \geqslant m \iff \gamma \in \Gamma_m \gamma_m^n. \end{aligned}$$

□

Preliminary results on natural boundaries.

LEMMA 4.13. *Let $h, \beta \in \mathbf{R}_{>0}$ with $\beta < 1$. Then the power series*

$$G_h(z) := \sum_{n\geqslant 1} |n|_p^h z^n, \qquad H_\beta(z) := \sum_{n\geqslant 1} \beta^{|n|_p^{-1}} z^n \tag{16}$$

have radius of convergence 1 and define holomorphic functions that have a natural boundary along the unit circle.

PROOF. That the radius of convergence is 1 follows from the fact that

$$\limsup_{n\to\infty} \left(|n|_p^h\right)^{1/n} = 1 = \limsup_{n\to\infty} \left(\beta^{|n|_p^{-1}}\right)^{1/n}.$$

Now, note that G_h and H_{β^p} satisfy the following similar functional equations:

$$\begin{aligned} G_h(z) &= \frac{z}{1-z} - \frac{z^p}{1-z^p} + p^{-h} G_h(z^p), \\ H_\beta(z) &= \beta\left(\frac{z}{1-z} - \frac{z^p}{1-z^p}\right) + H_{\beta^p}(z^p). \end{aligned} \tag{17}$$

In order to prove the statement on the natural boundary, we will show by induction on $k \geqslant 1$ that for every primitive p^k-th root of unity ω we have

$$\lim_{\lambda\to 1^-} G_h(\lambda\omega) = -\infty = \lim_{\lambda\to 1^-} H_\beta(\lambda\omega).$$

We present details for the case of $G_h(z)$; the proof for $H_\beta(z)$ is analogous. For $k = 1$, it follows from (17) that for every $0 < \lambda < 1$ we have

$$G_h(\lambda\omega) = \frac{\lambda\omega}{1-\lambda\omega} - \frac{\lambda^p}{1-\lambda^p} + p^{-h} G_h(\lambda^p).$$

As $G_h(\lambda^p) \leqslant \lambda^p/(1-\lambda^p)$ and $h > 0$, it follows that $G_h(\lambda\omega) \to -\infty$ as $\lambda \to 1^-$. For $k > 1$, the result follows from induction by substituting $\lambda\omega$ into (17). □

REMARK 4.14. Alternatively, since (17) implies that $G_h(z)$ is a so-called p-Mahler function, we could have immediately concluded that $G_h(z)$ is either rational or has the unit circle as a natural boundary by a result of Randé [**39**] (see also [**5**, Thm. 2]). The former possibility can be excluded by an explicit computation using the functional equation (17). Such an approach does not work for $H_\beta(z)$.

Proofs of Theorems 3.5 and 3.6. Assume that f satisfies **(H1)**–**(H3)**. We have already dealt with the case where f is coseparable in Proposition 4.7, so it remains to consider the case where f is not coseparable.

Using Lemma 3.1, we may write the zeta function $\zeta_f(z)$ as

$$\zeta_f(z) = \zeta_{f|_C}(z) \cdot \exp\left(\sum_{n\geqslant 1}\sum_{\gamma\in\Gamma} \frac{\deg(\sigma^n - \gamma)}{p^{v(\sigma^n-\gamma)}} \frac{z^n}{n}\right)^{1/|\Gamma|},$$

with a similar expression for the tame zeta function $\zeta_f^*(z)$. For $m \in \mathbf{Z}_{\geqslant 0}$, we consider separately the terms corresponding to a fixed value of $v(\sigma^n - \gamma)$, giving rise to functions

$$\zeta_{f,m}(z) = \exp\left(\sum_{\substack{n\geqslant 1, \gamma\in\Gamma \\ v(\sigma^n-\gamma)=m}} \frac{\deg(\sigma^n - \gamma)}{p^m} \frac{z^n}{n} \right).$$

Consider the sets

$$T_m = \{(n,\gamma) \in \mathbf{Z}_{>0} \times \Gamma \mid v(\sigma^n - \gamma) \geqslant m\}.$$

By Lemma 4.12, we have $T_m = \{(s_m n, \gamma\gamma_m^n) \mid n \in \mathbf{Z}_{>0}, \gamma \in \Gamma_m\}$, and hypothesis **(H3)** implies that the function

$$F_m(z) = \exp\left(\sum_{(n,\gamma)\in T_m} \deg(\sigma^n - \gamma) \frac{z^n}{n} \right)$$

satisfies $F_m(z)^{s_m/|\Gamma|} \in \mathbf{C}(z^{s_m})$; hence it is root-rational. It follows that the function

$$\zeta_{f,m}(z) = (F_m(z)/F_{m+1}(z))^{1/p^m}$$

is root-rational as well.

We analogously define the tame functions $\zeta_{f,m}^*(z)$, summing only over indices n coprime to p. By Equation (8), these are also root-rational, and we have the product formulas

$$\zeta_f(z) = \zeta_{f|_C}(z) \left(\prod_{m\geqslant 0} \zeta_{f,m}(z) \right)^{1/|\Gamma|} \quad \text{and} \quad \zeta_f^*(z) = \zeta_{f|_C}^*(z) \left(\prod_{m\geqslant 0} \zeta_{f,m}^*(z) \right)^{1/|\Gamma|}. \tag{18}$$

Our next aim is to simplify the tail (i.e. the product of all terms with m suitably large) of (18) using Proposition 4.3(v). To this end, take an integer $M \geqslant \max(N, v(p)/(p-1)+1)$ and set $r := s_M$ and $\tau := \sigma^{s_M}\gamma_M^{-1}$. By Lemma 4.10 the elements σ^{s_M} and γ_M commute, and we can rewrite the tail of (18) as

$$\prod_{m\geqslant M} \zeta_{f,m}(z) = \exp\left(\sum_{n\geqslant 1} \frac{\deg(\tau^n - 1)}{p^{v(\tau^n-1)}} \frac{z^{rn}}{rn} \right).$$

Set $C := v(\tau - 1) \geqslant M$. By Proposition 4.3(v) applied to $x = \tau$ and $y = 1$, we obtain

$$v(\tau^n - 1) = \begin{cases} C + v(n) & \text{if } \operatorname{char}(\operatorname{End}(G)) = 0; \\ C|n|_p^{-1} & \text{if } \operatorname{char}(\operatorname{End}(G)) = p > 0. \end{cases} \tag{19}$$

In particular, if $p \nmid n$, then $v(\tau^n - 1)$ is independent of n, so we have $\zeta_{f,m}^*(z) = 1$ for $m \geqslant M$ and $m \neq C$. The product expansion in (18) therefore shows that the tame zeta function is root-rational, proving Theorem 3.6.

Now suppose that f also satisfies **(H4)**. Since s divides r and $\deg(\tau^n - 1) = \deg(\sigma^{rn} - \tilde{\gamma}^{rn/s})$, we see that $(\deg(\tau^n - 1))_{n\geqslant 1}$ is a linear recurrent sequence with a unique dominant root, say Λ, with multiplicity $\mu \in \mathbf{Z}$. We then obtain

$$z\frac{d}{dz}\log \prod_{m\geqslant M} \zeta_{f,m}(z) = \sum_{n\geqslant 1} \mu\Lambda^n z^{rn} \cdot \begin{cases} p^{-C}|n|_p^{v(p)} & \text{if } \operatorname{char}(\operatorname{End}(G)) = 0 \\ p^{-C|n|_p^{-1}} & \text{if } \operatorname{char}(\operatorname{End}(G)) = p \end{cases} + R(z),$$

where $R(z)$ is some power series with radius of convergence $> |\Lambda|^{-1/r}$.

As stated in [**6**, Lemma 1], the existence of a natural boundary for a series $\sum a_n z^n$ along its circle of convergence implies the existence of a natural boundary along its circle of convergence for the corresponding zeta function $\exp \sum a_n z^n/n$. By applying Lemma 4.13 with $h = v(p)$ or $\beta = p^{-C}$ (depending on whether $\operatorname{End}(G)$ is of characteristic 0 or p) and substituting Λz^r for z into $G_h(z)$ or $H_\beta(z)$, it follows that the series

$$\left(\prod_{m\geqslant M} \zeta_{f,m}(z)\right)^{1/|\Gamma|}$$

has a natural boundary along its circle of convergence. Theorem 3.5 now follows from the product expansion in (18). □

Remark 4.15. An examination of the (by the proof, finite) product expansion for the tame zeta function (18), shows that in fact $\zeta_f^*(z)^t \in \mathbf{C}(z)$ for $t = p^{C+1}r$, where C and r are as in the proof.

5. Discussion of the hypotheses

Classification of G for $V = \mathbf{P}^1$. Suppose that $V = \mathbf{P}^1$ and recall that K is algebraically closed. Since Γ is finite and ι has Zariski-dense image, for dimension reasons G is a connected one-dimensional algebraic group. By the Barsotti–Chevalley structure theorem for algebraic groups [**13, 14**], G is an extension of a linear algebraic group by an abelian variety, and thus (again by connectedness and dimension considerations) G is either a one-dimensional connected linear algebraic group or an abelian variety of dimension one. In the latter case, G is an elliptic curve E. In the former case, either $G = \mathbf{G}_m$, the multiplicative group; or $G = \mathbf{G}_a$, the additive group [**47**, Thm. 3.4.9]. We denote by $[m]$ the multiplication-by-m map on G. The corresponding endomorphism rings are as follows:

(i) if $G = \mathbf{G}_m$, then $\operatorname{End}(G) \cong \mathbf{Z}$, with the map $[m]$ given by $x \mapsto x^m$;
(ii) if $G = \mathbf{G}_a$, then $\operatorname{End}(G) \cong K\langle\phi\rangle$ is the ring of skew-commutative polynomials in the Frobenius $\phi\colon x \mapsto x^p$, with $\phi a = a^p\phi$ for all $a \in K$;
(iii) if $G = E$ is an elliptic curve, then $\operatorname{End}(E)$ is either $\mathbf{Z}$, an order in an imaginary quadratic number field in which p splits, or a maximal order in the quaternion algebra over $\mathbf{Q}$ that ramifies precisely at p and ∞ [**17**].

Hypothesis (H1).

Lemma 5.1. *If $f\colon C \to C$ is an arbitrary map on a finite set C, then $\zeta_f(z)$ is rational and $\zeta_f^*(z)$ is root-rational.*

Proof. Since there are only finitely many orbits, this follows for $\zeta_f(z)$ from the Euler product (3), and then for $\zeta_f^*(z)$ from (5). □

Corollary 5.2. *If $f\colon V \to V$ is a dynamically affine map and either* $\dim V = 1$ (*e.g. if* $V = \mathbf{P}^1$) *or G is complete* (*e.g. if G is an abelian variety*), *then f satisfies* **(H1)**.

PROOF. If G is complete, then $C = \emptyset$, and the result is clear. If $\dim V = 1$, then the assumption that $\Gamma\backslash G$ is a Zariski-dense open subset of V implies that C is finite, and Lemma 5.1 applies. □

EXAMPLE 5.3. We give an example where **(H1)** fails. Consider $G = \mathbf{G}_m \times \mathbf{G}_m$, $\Gamma = \{1\}$, the standard embedding $G \hookrightarrow V := \mathbf{P}^1 \times \mathbf{P}^1$ and

$$f\colon V \to V,\ (x, y) \mapsto (x^m, y^m)$$

for an integer $m \geqslant 2$ coprime to p. Then

$$C = (\mathbf{P}^1 \times \{0\}) \cup (\mathbf{P}^1 \times \{\infty\}) \cup (\{0\} \times \mathbf{P}^1) \cup (\{\infty\} \times \mathbf{P}^1)$$

is a union of four copies of $\mathbf{P}^1$ intersecting in four points

$$C' = \{(0, 0), (0, \infty), (\infty, 0), (\infty, \infty)\},$$

all of which are fixed by f. Applying Remark 2.2, we get that

$$\zeta_{f|_C}(z) = \zeta_g(z)^4(1 - z)^4,$$

where $\zeta_g(z)$ is the zeta function of $g\colon \mathbf{P}^1 \to \mathbf{P}^1$, $x \mapsto x^m$, which was shown to be transcendental over $\mathbf{C}(z)$ by Bridy [**8**] (our result in this paper even shows that the function has a natural boundary).

Hypothesis (H2).

PROPOSITION 5.4.

(i) *A nontrivial abelian variety A satisfying* **(H2)** *with $\mathscr{R} = \mathrm{End}(A)$ has to be simple.*

(ii) *There exist commutative algebraic groups G of arbitrary dimension > 1 that satisfy* **(H2)** *with $\mathscr{R} = \mathrm{End}(G)$ but that are not simple.*

(iii) *For any connected commutative algebraic group G,* **(H2)** *is equivalent to the claim that all nonzero elements of $\mathscr{R}$ are isogenies and for every m, the set*

$$I_m := \{\tau \in \mathscr{R} - \{0\} \mid \log_p \deg_{\mathrm{i}}(\tau) \geqslant m\} \cup \{0\}$$

is an ideal in $\mathscr{R}$. (Note that by the multiplicativity of the inseparable degree this is equivalent to $\deg_{\mathrm{i}}(\tau_1 + \tau_2) \geqslant \min\{\deg_{\mathrm{i}}(\tau_1), \deg_{\mathrm{i}}(\tau_2)\}$ *for all nonzero τ_1, τ_2, $\tau_1 \neq -\tau_2$.)*

(iv) *Let $G = A$ be a nontrivial abelian variety. Then* **(H2)** *holds with* $\mathscr{R} = \mathbf{Z} \hookrightarrow \mathrm{End}(A)$ *(and then necessarily $\Gamma = \{1\}$ or $\Gamma = \{\pm 1\}$).*

PROOF. Note that the hypothesis on the existence of v implies that $\mathscr{R}$ is a (not necessarily commutative) domain (Proposition 4.3(i)).

(i) Since an abelian variety A factors up to isogeny into a direct product of simple abelian varieties, $\mathrm{End}(A)$ is a domain if and only if A is simple.

(ii) Consider extensions of algebraic groups

$$1 \to \mathbf{G}_m \to G \to A \to 1,$$

where G is abelian and A is any simple abelian variety. These are classified by $\mathrm{Ext}^1(A, \mathbf{G}_m) \cong \widehat{A}(K)$, where $\widehat{A}$ is the dual abelian variety of A [**33**, Thm. 9.3]. Suppose $\widehat{A}(K)$ has a non-torsion point P (in particular, K has to be transcendental over $\mathbf{F}_p$) and choose an extension corresponding to P. We claim that G does not contain any nontrivial abelian variety. Suppose otherwise and let A' be a nontrivial abelian variety contained in G. The image of A' in A cannot be zero, and hence is equal to all of A since A is simple. It follows that A' and $\mathbf{G}_m$ generate G, and a result of Arima [**1**, Thm. 2] implies that the extension corresponds to a point of finite order in $\widehat{A}(K)$. We conclude that G does not contain any nontrivial abelian variety, and hence by [**1**, Prop. 7] the restriction map $\mathrm{End}(G) \to \mathrm{End}(\mathbf{G}_m) \cong \mathbf{Z}$ is injective, meaning that $\mathrm{End}(G) \cong \mathbf{Z}$. The inseparable degree of $[n]$ on G is the product of those on $\mathbf{G}_m$ and A, and if A has dimension g and p-rank r, then the valuation $v = (2g - r + 1)v_p$ on $\mathscr{R} = \mathrm{End}(G)$ satisfies **(H2)**.

(iii) If v is a valuation as in **(H2)**, then

$$I_m = \{\tau \in \mathscr{R} \mid v(\tau) \geqslant m\}$$

is an ideal. Conversely, if all I_m are ideals, then v defined by

$$v(\tau) := \sup\{m \mid \tau \in I_m\}$$

satisfies **(H2)**.

(iv) For a nontrivial abelian variety A the endomorphism ring $\mathrm{End}(A)$ has characteristic zero and the maps $[m]\colon A \to A$ for $m \in \mathbf{Z} \setminus \{0\}$ are isogenies, and are separable if and only if $p \nmid m$. By multiplicativity of the inseparable degree, we find that $\deg_{\mathrm{i}}([m]) = \deg_{\mathrm{i}}([p])^{v_p(m)}$, and hence the valuation $v\colon \mathscr{R} \to \mathbf{Z} \cup \{\infty\}$ given by $v([m]) = cv_p(m)$ with $c := \log_p \deg_{\mathrm{i}}([p])$ satisfies **(H2)**. ☐

Lemma 5.5. *Hypothesis* **(H2)** *holds for dynamically affine maps on* $\mathbf{P}^1$.

Proof. We verify, for G a one-dimensional connected algebraic group, that the I_m as in Proposition 5.4(iii) are indeed ideals. Note that the claim that all nonzero elements of $\mathrm{End}(G)$ are isogenies is immediate by a dimension argument. For $G = \mathbf{G}_m$ and $G = \mathbf{G}_a$, the set of inseparable isogenies together with the zero map is the principal ideal generated by the Frobenius $\phi\colon x \mapsto x^p$, so $I_m = (\phi^m)$ is an ideal. If $G = E$ is an elliptic curve, then for any isogeny $\tau\colon E \to E$, we have that $\log_p \deg_{\mathrm{i}}(\tau)$ is the largest $r > 0$ for which τ factors through the p^r-Frobenius $E \to E^{(p^r)}$ [**44**, II.2.12], which again implies that the I_m are ideals. ☐

Remark 5.6. Another approach, folllowing [**9**], is to check the result for each of the possible one-dimensional groups G with the following valuations:

(i) if $G = \mathbf{G}_m$, then on $\mathrm{End}(G) \cong \mathbf{Z}$ set $v = v_p$, the p-adic valuation;
(ii) If $G = \mathbf{G}_a$, then on $\mathrm{End}(G) = K\langle\phi\rangle$ set $v = v_\phi$, the valuation associated to the two-sided ideal (ϕ);
(iii) If $G = E$ is an elliptic curve, set $v = v_p \circ N$, where N is the field norm of the extension $\mathrm{End}(E) \otimes \mathbf{Q}$ of $\mathbf{Q}$ if E is ordinary and N is the reduced norm on the quaternion algebra $\mathrm{End}(E) \otimes \mathbf{Q}$ if E is supersingular.

Hypothesis (H3). The following general observation will be used multiple times to verify that Hypothesis **(H3)** holds in certain cases: *If R is a (not necessarily commutative) domain and Γ is a nontrivial finite subgroup of the multiplicative group of R, then $\sum_{\gamma \in \Gamma} \gamma = 0$.*

We first discuss the degree function on a commutative subring $\mathscr{R}$ of the endomorphism ring $\mathrm{End}(A)$ of an abelian variety A. The ring $S = \mathrm{End}(A) \otimes_{\mathbf{Z}} \overline{\mathbf{Q}}$ is a semisimple $\overline{\mathbf{Q}}$-algebra, and hence is isomorphic to a product of finitely many (full) rings of matrices over $\overline{\mathbf{Q}}$. Let

$$\psi = (\psi_1, \dots, \psi_k)\colon S \to \prod_{i=1}^{k} \mathrm{M}_{n_i}(\overline{\mathbf{Q}})$$

be such an isomorphism. The degree of an endomorphism $\alpha \in \mathrm{End}(A)$ can be computed by the formula

$$\deg(\alpha) = \prod_{i=1}^{k} \det \psi_i(\alpha)^{\nu_i},$$

where $\nu_i \in \mathbf{Z}_{>0}$ are certain integers (see [**25**, Cor. 3.6] or the discussion in [**11**, Prop. 2.3]). Since the ring $\mathscr{R}$ is commutative, the matrices in $\psi_i(\mathscr{R}) \subseteq \mathrm{M}_{n_i}(\overline{\mathbf{Q}})$ can be simultaneously triangularised, so that after conjugating by appropriate matrices, we may assume that $\psi(\mathscr{R})$ lies in the product of rings $\mathrm{UT}_{n_i}(\overline{\mathbf{Q}})$ of upper triangular matrices. Composing the homomorphism $\psi_i|_{\mathscr{R}}$ with the homomorphism $\mathrm{UT}_{n_i}(\overline{\mathbf{Q}}) \to \overline{\mathbf{Q}}^{n_i}$ that maps each matrix to the tuple consisting of its diagonal elements, we obtain a ring homomorphism

$$\lambda = (\lambda_1, \ldots, \lambda_l) \colon \mathscr{R} \to \overline{\mathbf{Q}}^{l},$$

where $l = \sum_{i=1}^{k} n_i$. The degree function on $\mathscr{R}$ then takes the form

$$\deg(\alpha) = \prod_{j=1}^{l} \lambda_j(\alpha)^{\mu_j},$$

where $\mu_j \in \mathbf{Z}_{>0}$ are certain integers.

PROPOSITION 5.7. *Let A be an abelian variety over K, let f be a dynamically affine map with $G = A$, and let $\mathscr{R}$ be a commutative subring of the endomorphism ring* $\mathrm{End}(A)$ *that contains both σ and Γ. If* **(H2)** *is satisfied for the ring $\mathscr{R}$, then f satisfies* **(H3)**.

PROOF. Using the notation provided by the statement of **(H3)**, write $\tau := \sigma^{s_m}$ and $\tilde{\tau} := \tau\gamma_m^{-1}$. By Lemma 4.1(ii), confinedness of σ implies that $\tau^n - \gamma$ is an isogeny for all $\gamma \in \Gamma$. Since $\mathscr{R}$ is commutative, we have $\deg(\tau^n - \gamma\gamma_m^n) = \deg(\tilde{\tau}^n - \gamma)$ for $\gamma \in \Gamma$.

Using the notation explained at the beginning of this subsection, we obtain the formula

$$\sum_{\gamma \in \Gamma_m} \deg(\tilde{\tau}^n - \gamma) = \sum_{\gamma \in \Gamma_m} \prod_{j=1}^{l} \lambda_j(\tilde{\tau}^n - \gamma)^{\mu_j}.$$

Since $\lambda_j \colon \mathscr{R} \to \overline{\mathbf{Q}}$ are ring homomorphisms, we may expand the product on the right hand side and rewrite the formula as

$$\sum_{\gamma \in \Gamma_m} \prod_{j=1}^{l} \lambda_j(\tilde{\tau}^n - \gamma)^{\mu_j} = \sum_{k} \sum_{\gamma \in \Gamma_m} \chi_k(\gamma)\eta_k^n$$

for some $\eta_k \in \overline{\mathbf{Q}}$ and some characters $\chi_k \colon \Gamma \to \overline{\mathbf{Q}}^{\times}$. If a character χ_k is nontrivial, then $\sum_{\gamma \in \Gamma_m} \chi_k(\gamma) = 0$; otherwise, $\sum_{\gamma \in \Gamma_m} \chi_k(\gamma) = |\Gamma_m|$. Thus, we obtain

$$\frac{1}{|\Gamma_m|} \sum_{\gamma \in \Gamma_m} \deg(\tau^n - \gamma\gamma_m^n) = \sum_{\chi_k = 1} \eta_k^n,$$

and the desired result follows from the general criteria for rationality of the zeta function discussed in Section 2. □

PROPOSITION 5.8. *A dynamically affine map on $\mathbf{P}^1$ satisfies* **(H3)**.

Proof. As in the proof of the previous proposition, we set $\tau := \sigma^{s_m}$. If $G = \mathbf{G}_m$, then, identifying $\mathrm{End}(G)$ with $\mathbf{Z}$, we have

$$\deg(\tau^n - \gamma\gamma_m^n) = |\tau^n - \gamma\gamma_m^n| = \deg(\tau)^n - \gamma\gamma_m^n \operatorname{sgn}(\tau)^n.$$

If $G = \mathbf{G}_a$, then

$$\deg(\tau^n - \gamma\gamma_m^n) = \deg(\tau)^n.$$

(This holds even when $\deg(\tau) = 1$ since τ is confined.) Finally, if $G = E$ is an elliptic curve, then $\deg(\tau) = \tau\overline{\tau}$, where $\overline{\tau}$ denotes τ or the complex/quaternionic conjugate of τ depending on whether $\mathrm{End}(E)$ is $\mathbf{Z}$, an order in an imaginary quadratic number field or an order in a quaternion algebra over $\mathbf{Q}$. In either case

$$\begin{aligned}\sum_{\gamma\in\Gamma_m} \deg(\tau^n - \gamma\gamma_m^n) &= \sum_{\gamma\in\Gamma_m} \left((\tau\overline{\tau})^n - \gamma\gamma_m^n\overline{\tau}^n - \tau^n\overline{\gamma_m}^n\overline{\gamma} + 1\right)\\ &= |\Gamma_m|(\deg(\tau)^n + 1) - \left(\sum_{\gamma\in\Gamma_m}\gamma\right)\gamma_m^n\overline{\tau}^n - \tau^n\overline{\gamma_m}^n\left(\sum_{\gamma\in\Gamma_m}\overline{\gamma}\right).\end{aligned}$$

Combining the three cases, we find that in case Γ_m is nontrivial, we get

$$\sum_{\gamma\in\Gamma_m} \deg(\tau^n - \gamma\gamma_m^n) = \begin{cases} |\Gamma_m| \deg(\tau)^n & \text{if } G = \mathbf{G}_m \text{ or } G = \mathbf{G}_a;\\ |\Gamma_m|(\deg(\tau)^n + 1) & \text{if } G = E, \end{cases} \tag{20}$$

whereas in case Γ_m is trivial, the elements τ and γ_m commute by Lemma 4.10, and hence

$$\begin{aligned}\sum_{\gamma\in\Gamma_m} \deg(\tau^n - \gamma\gamma_m^n) &= \deg(\tau^n - \gamma_m^n)\\ &= \begin{cases} \deg(\tau)^n - (\gamma_m \operatorname{sgn}(\tau))^n & \text{if } G = \mathbf{G}_m;\\ \deg(\tau)^n & \text{if } G = \mathbf{G}_a;\\ \deg(\tau)^n + 1 - (\gamma_m\overline{\tau})^n - (\tau\overline{\gamma_m})^n & \text{if } G = E.\end{cases}\end{aligned} \tag{21}$$

We may regard these formulas as equalities between complex numbers (for $G = E$ embedding the field $\mathbf{Q}(\gamma_m\overline{\tau})$ into $\mathbf{C}$). It follows that the corresponding zeta function is rational. □

Example 5.9. For an example where **(H3)** does not hold, consider $K = \overline{\mathbf{F}}_3$, $G = \mathbf{G}_a \times \mathbf{G}_a$, $\Gamma = \{1\}$, $V = G$ and

$$f\colon G \to G, (x, y) \mapsto (x^9 + y^3, x^3).$$

Since the differential of $\sigma = f$ is zero, the map f is even coseparable. One may directly compute the values of $\deg(\sigma^n - 1)$. One way to do this is to write

$$\sigma = \begin{pmatrix} \phi^2 & \phi \\ \phi & 0 \end{pmatrix}$$

with $\phi\colon \mathbf{G}_a \to \mathbf{G}_a$ the Frobenius map, and show that for a matrix τ in $\mathrm{M}_2(\mathbf{F}_3[\phi])$ with nonzero determinant the degree of τ as a map $\tau\colon G \to G$ and the degree in ϕ of $\det(\tau) \in \mathbf{F}_3[\phi]$ are related by the formula

$$\deg(\tau) = 3^{\deg_\phi(\det(\tau))}.$$

(This follows easily by writing τ in Smith normal form.) Computing the eigenvalues of σ as Laurent series in ϕ^{-1}, we get that

$$\deg(\sigma^n - 1) = \begin{cases} 9^n, & 2\nmid n; \\ 9^{n-|n|_3^{-1}}, & 2|n, \end{cases}$$

and hence the zeta function satisfies the equation

$$z\frac{d}{dz}\log\zeta_f(z) = \sum \deg(\sigma^n - 1)z^n = \frac{9z}{1-81z^2} + H_{1/9}(81z^2),$$

where $H_{1/9}(z)$ is the function from Lemma 4.13. It follows that $\zeta_f(z)$ has a natural boundary along $|z| = 1/9$ and **(H3)** indeed fails to hold.

For a detailed computation of the degree and a general discussion of fixed points of endomorphisms of vector groups, we refer the reader to [**12**].

Hypothesis (H4).

PROPOSITION 5.10. *A dynamically affine non-coseparable map on* $\mathbf{P}^1$ *satisfies* **(H4)**.

PROOF. It follows directly from (20) and (21) applied for $m = N$ that $\deg(\tau)$ is the dominant root (note that for $G = \mathbf{G}_m$ and $G = E$, confinedness of σ implies that $\deg(\tau) \geqslant 2$). ☐

We will now examine property **(H4)** in the case where $G = A$ is an abelian variety.

PROPOSITION 5.11. *Let f be a dynamically affine map with $G = A$ an abelian variety. Assume that the hypothesis* **(H2)** *is satisfied for a commutative ring* $\mathscr{R} \subseteq \mathrm{End}(A)$.

(i) *The map f satisfies* **(H4)** *if and only if σ is not coseparable and the characteristic polynomial of the action of σ on the ℓ-adic Tate module $T_\ell(A)$, where ℓ is any prime $\ell \neq p$, has no roots of complex absolute value 1.*

(ii) *The map f satisfies* **(H4)** *if and only if the map σ, regarded as a dynamically affine map $\sigma\colon A \to A$, satisfies* **(H4)**.

PROOF. By Lemma 4.11 it suffices to treat the case where σ is non-coseparable and the number s exists.

(i) We have the formula

$$\deg(\sigma^{sn} - \tilde{\gamma}^n) = \prod_{j=1}^{l} (\lambda_j(\sigma)^{sn} - \lambda_j(\tilde{\gamma})^n)^{\mu_j}\,. \tag{22}$$

Since Γ is finite and σ is confined, the elements $\lambda_j(\tilde{\gamma})$ are roots of unity while $\lambda_j(\sigma)$ are not. By the discussion in [**11**, Section 5], the elements $\lambda_j(\sigma)$ are exactly the roots of the action of σ on $T_\ell(A)$. Expanding the expression in (22), one can show that $\deg(\sigma^{sn} - \tilde{\gamma}^n)$ is given by a linear recurrence, and that the dominant root is unique if and only if $|\lambda_j(\sigma)| \neq 1$ for all j. (The argument is identical to the one where $\tilde{\gamma} = 1$ as given in [**11**, Prop. 5.1(v)].)

(ii) This is clear since the condition in (i) depends neither on s nor on $\tilde{\gamma}$. ☐

Example 5.12. For an example where **(H4)** fails, let $G = \mathbf{G}_m^4$, $\Gamma = \{1\}$ and $V = G$. We choose $f = \sigma \in \operatorname{End}(G) \cong \mathrm{M}_4(\mathbf{Z})$ to be the companion matrix of the minimal polynomial g of a Salem number $\alpha > 1$ of degree 4 (e.g. $g = x^4 - 3x^3 + 3x^2 - 3x + 1$ [**46**]). Then $\deg(\sigma^n - 1) = |\det(\sigma^n - 1)|$ and if $\beta \in \mathbf{C}$ is a zero of g with absolute value 1, then α and $\alpha\beta$ are distinct dominant roots of the linear recurrent sequence $(\deg(\sigma^n - 1))_{n \geqslant 1}$.

Proofs of Theorems A and B.

Proof of Theorem A. The theorem will follow by combining Theorems 3.5 and 3.6 with the following observations.

A dynamically affine map $f\colon \mathbf{P}^1 \to \mathbf{P}^1$ satisfies the hypotheses **(H1)**–**(H3)** by Corollary 5.2, Lemma 5.5 and Proposition 5.8. If f is not coseparable, it satisfies **(H4)** by Proposition 5.10. If f is coseparable, the function $\zeta_f(z)$ is rational by Proposition 4.7. □

Proof of Theorem B. In this situation, **(H1)**–**(H3)** hold by Corollary 5.2 and Propositions 5.4(iv) and 5.7. If $p|m$, the map σ is coseparable, and $\zeta_f(z)$ is rational. If $p\nmid m$, σ is not coseparable, and f satisfies **(H4)** by Proposition 5.11. □

Appendix A.
Radius of convergence of ζ_f for dynamically affine maps f

In general, the existence of a positive radius of convergence of a dynamical zeta function is a nontrivial property of the growth of the number of periodic points of a given order. In the manifold setting, this is studied in [**2**]; Kaloshin showed that a positive convergence radius is *not* topologically Baire generic in the C^r topology for any $2 \leqslant r < +\infty$ [**30**, Corollary 1].

In this appendix, we study this problem for a morphism $f\colon V \to V$ on an algebraic variety V. We can say something in case f is dynamically affine, or in case V is smooth projective, but we do not know what happens in the general case.

Theorem A.1. *Let $f\colon V \to V$ denote a dynamically affine map over an algebraically closed field K of characteristic p, satisfying* **(H1)**. *Then the zeta functions $\zeta_f(z)$ and $\zeta_f^*(z)$ converge to holomorphic functions on a nontrivial open disk centred at the origin.*

Proof. It follows from the definitions that $\zeta_f^*(z)$ converges whenever $\zeta_f(z)$ does. The latter function converges for $|z| < 1/\limsup \sqrt[n]{f_n}$. Hence to prove the statement, it suffices to prove that $f_n \leqslant c^n$ for some constant c. By Lemma 3.1, it suffices to prove that $(f|_C)_n \leqslant c^n$ and $\#\ker(\sigma^n - \gamma) \leqslant c^n$ for all $\gamma \in \Gamma$ and some constant c (independent of n). The first statement follows immediately from **(H1)**.

For the second statement, we note that $\#\ker(\sigma^n - \gamma) = \#\mathrm{Fix}(\tau)$, where $\tau = \sigma^n\gamma^{-1}$. The main point in the proof is to reduce to the case of G being an abelian variety, a torus, or a vector group, by a method similar to the one employed in [**12**]. Here, we give a more ad hoc discussion (avoiding cohomology) and simplify matters using the commutativity of G.

We first observe that if N is a connected normal algebraic subgroup of G stable under $\operatorname{End}(G)$, then τ induces an endomorphism $\tau_{G/N}$ of G/N. We claim that $\tau_{G/N}$ is confined and that

$$\#\mathrm{Fix}(\tau) = \#\mathrm{Fix}(\tau|_N) \cdot \#\mathrm{Fix}(\tau_{G/N}). \tag{23}$$

To see this, we first note that by Lemma 4.1(i) powers τ^k of τ are of the form $\gamma'\sigma^{nk}$ for some $\gamma' \in \Gamma$, and hence by Lemma 4.1(ii) τ is confined. Since N is connected and the map τ is confined, we get that $\tau|_N - 1$ is an isogeny (in particular, it is surjective), which implies that the map $\mathrm{Fix}(\tau) \to \mathrm{Fix}(\tau_{G/N})$ is surjective as well. Applying this to $\gamma = 1$ shows that the map $\sigma_{G/N}$ is confined, and we get an exact sequence of finite groups

$$0 \to \mathrm{Fix}(\tau|_N) \to \mathrm{Fix}(\tau) \to \mathrm{Fix}(\tau_{G/N}) \to 0.$$

Notice also that $\tau|_N = \sigma|_N^n \gamma|_N^{-1}$ and $\tau_{G/N} = \sigma_{G/N}^n \gamma_{G/N}^{-1}$ admit the same decomposition as τ with $\sigma|_N$ (resp., $\sigma_{G/N}$) being a confined isogeny on N (resp., G/N).

We apply (23) several times: first, by Chevalley's structure theorem for algebraic groups [**14**, Thm. 1.1], G has a unique normal connected linear algebraic subgroup N such that G/N is an abelian variety. Then N is stable by $\mathrm{End}(G)$, since there are no nontrivial morphisms from a linear algebraic group N to an abelian variety A [**14**, Lem. 2.3]. Now suppose G is a connected commutative linear algebraic group; then there exists a normal connected unipotent algebraic subgroup U of R such that the quotient R/U is a torus T, i.e. isomorphic to $\mathbf{G}_m^s$ for some $s \in \mathbf{Z}_{\geqslant 0}$ [**34**, Thm. 16.33]. There are no nontrivial morphisms $U \to T$ [**16**, Cor. IV.2.2.4], so U is preserved by any endomorphism of R. Now if G is connected commutative unipotent, it is isogenous to a direct product $W_1 \times \cdots \times W_t$ of additive groups of truncated Witt vectors [**42**, Thm. VII.1]. Since $p^d W_i = 0$ for some d, we obtain a decomposition series of G (using [**16**, Prop. IV.2.2.3]) $G \supseteq pG \supseteq p^2G \supseteq \cdots \supseteq 0$, in which pG is preserved by any endomorphism of G, and each successive quotient is a connected commutative unipotent algebraic group of exponent p. By [**42**, Prop. VII.11], such a group is isomorphic to a vector group $\mathbf{G}_a^r$ for some $r \in \mathbf{Z}_{\geqslant 0}$.

By the above discussion, we are reduced to considering the following three cases. In each of these cases, G is connected commutative, $\mathrm{End}(G)$ is a ring with a degree function $\deg\colon \mathrm{End}(G) \to \mathbf{N} \cup \{-\infty\}$ and $\#\ker(\sigma^n - \gamma) \leqslant \deg(\sigma^n - \gamma)$, so it suffices to prove that in each of these cases $\deg(\sigma^n - \gamma)$ grows at most exponentially in n.

- G **is an abelian variety**: G is isogenous to a product of simple abelian varieties, and $\deg(\sigma^n - \gamma)$ becomes a product of reduced norms $N(\sigma_i^n - \gamma_i)$ on finitely many simple $\mathbf{Q}$-algebras R_i (with $\tau_i \in R_i$ and $\gamma_i \in R_i^\times$) [**11**, Prop. 2.3]. Passing to the algebraic closure of $\mathbf{Q}$, one easily sees that these satisfy a linear recurrence in n, and hence grow at most exponentially.
- $G \simeq \mathbf{G}_m^s$ **is a torus**: Identifying endomorphisms of G with matrices in $\mathrm{M}_s(\mathbf{Z})$, one sees (e.g. by using the Smith normal form) that

$$\deg(\sigma^n - \gamma) = |\det(\sigma^n - \gamma)|.$$

  Expanding the determinant shows the desired growth behaviour.
- $G \simeq \mathbf{G}_a^r$ **is a vector group**: Endomorphisms of G are given by $r \times r$ matrices over the skew polynomial ring $K\langle\phi\rangle$ with $\phi a = a^p \phi$ for $a \in K$. The degree of an isogeny $\tau \in \mathrm{End}(G)$ can be computed using the Dieudonné determinant ddet by the formula

$$\deg(\tau) = p^{\deg_\phi \mathrm{ddet}(\tau)}.$$

(Since $K\langle\phi\rangle$ is left and right euclidean, we can put the matrix τ in Smith normal form and use the fact that unimodular matrices have Dieudonné determinant of degree 0 [**24**, Thm. 4.6]). We will use that if τ is a matrix over $K\langle\phi\rangle$ all of whose entries have degree $\leqslant d$ as polynomials in ϕ, then $\deg_\phi \operatorname{ddet}(\tau) \leqslant rd$ [**24**, Thm. 3.5]. Choose an integer $d \geqslant 1$ so that the entries of σ have degree $\leqslant d$. For sufficiently large n the entries of $\sigma^n - \gamma$ have degree $\leqslant nd$, and hence

$$\deg(\sigma^n - \gamma) = p^{\deg_\phi \operatorname{ddet}(\sigma^n - \gamma)} \leqslant p^{nrd}. \qquad \square$$

REMARK A.2. For a more comprehensive treatment of degrees and inseparable degrees of endomorphisms of algebraic groups (not necessarily commutative), we refer the reader to [**12**].

In the above proof, the positive radius of convergence of $\zeta_f(z)$ is recursively defined based on a decomposition of G along subgroups and quotients. The following is a case where we can find a direct bound on the radius of convergence:

PROPOSITION A.3. *When V is* smooth projective *and $f\colon V \to V$ is any morphism, $\zeta_f(z)$ and $\zeta_f^*(z)$ define holomorphic functions on a disk of radius the smallest absolute value of a zero of the characteristic polynomial*

$$\det(1 - f^*z \mid \mathrm{H}^{2\bullet}(V)),$$

where $\mathrm{H}^{2\bullet}(V) = \bigoplus_{j=0}^{\dim V} \mathrm{H}^{2j}(V, \mathbf{Q}_\ell)$ *is the even étale cohomology of V for some $\ell \neq p$.*

PROOF. In this case, we have a coefficient-wise bound $f_n \leqslant (\Gamma_{f^n} \cdot \Delta)$, where the right hand side is the intersection number of the graph of f^n with the diagonal in $V \times V$. Since V is smooth projective and f has finitely many fixed points, by the Grothendieck–Lefschetz fixed point formula in ℓ-adic cohomology for $\ell \neq p$ [**15**, Cor. 3.7, p. 152 (= Exposé "Cycle", p. 24)] we find that

$$\exp\Big(\sum (\Gamma_{f^n} \cdot \Delta) z^n/n\Big) = \prod_{i=0}^{2\dim V} \det(1 - f^*z \mid \mathrm{H}^i(V, \mathbf{Q}_\ell))^{(-1)^{i+1}}, \tag{24}$$

and the right hand side converges in an open disk of radius the smallest absolute value of a zero of the denominator. $\square$

REMARK A.4. If $V = \mathbf{P}^k$ is a projective space, the result follows essentially from Bézout's theorem (see e.g. [**18**, Prop. 1.3]). A more general "intersection-theoretic" argument such as in the proof of Proposition A.3 seems to apply only in a restrictive setting, since the Grothendieck–Lefschetz fixed point formula can fail for general endomorphisms of general varieties, and one cannot in general leave out the assumptions of properness and smoothness. It seems these assumptions are rarely satisfied, as witnessed by the following sample result in characteristic zero: if Γ is a finite group of endomorphisms of a complex abelian variety of dimension $\geqslant 3$ acting irreducibly on the tangent space at 0, $\Gamma\backslash G$ is necessarily a projective space, G is a power of an elliptic curve, and Γ is one of two possible groups (depending on the dimension) [**3**, Thm. 1.1].

Appendix B.
Explicit computation of tame zeta functions for some dynamically affine maps on $\mathbf{P}^1$

Classification of dynamically affine maps on $\mathbf{P}^1$. Bridy [**9**] classified all dynamically affine maps f of degree $\geqslant 2$ on the projective line by showing that they are conjugate by a fractional linear transformation to polynomials f_c in an explicit standard form, as in Table 1 (where $\mu_d \subseteq k^\times$ denotes a nontrivial subgroup consisting of d-th roots of unity). This is the characteristic-p analogue of the classification over $\mathbf{C}$ given in [**35**, Thm. 3.1].

G	Γ	$\Gamma\backslash G$	f_c
$\mathbf{G}_a$	$\{1\}$	$\mathbf{P}^1 - \{\infty\}$	Additive polynomial
	μ_d		Subadditive polynomial
$\mathbf{G}_m$	$\{1\}$	$\mathbf{P}^1 - \{0,\infty\}$	Power map
	$\{\pm 1\}$	$\mathbf{P}^1 - \{\infty\}$	Chebyshev map
E	$\neq \{1\}$	$\mathbf{P}^1$	Lattès map

TABLE 1. Classification of dynamically affine maps on $\mathbf{P}^1$.

With the notation and terminology of Table 1, f is coseparable precisely when either f_c is inseparable, or f_c is a separable (sub)additive polynomial for which $f_c'(0)$ is transcendental over k (cf. [**9**, Thm. 1.2 & 1.3]). One easily checks that these are precisely the maps for which f_n is *maximal* for all n (i.e. $f_n = \deg(f)^n + 1$; each fixed point of f^n has multiplicity one).

Some examples of tame zeta functions. The "trivial" case provides us with a useful notational tool: if $X = \mathrm{pt}$ is a point, then f has a unique fixed point ($f_n = 1$ for all n), so we suppress the (irrelevant) f from the notation, to obtain

$$\zeta_{\mathrm{pt}}(z) = \frac{1}{1-z} \qquad \text{and} \qquad \zeta^*_{\mathrm{pt}}(z) = \zeta_{\mathrm{pt}}(z)/\sqrt[p]{\zeta_{\mathrm{pt}}(z^p)} = \frac{\sqrt[p]{1-z^p}}{1-z}.$$

We will now present examples of tame zeta functions, writing them in a concise form using the function $\zeta^*_{\mathrm{pt}}(az^b)$ for various a and b. Much of the general structure of (tame) zeta functions of algebraic groups is already visible in the following basic example for which we provide a detailed computation (we stick to $p > 2$ for convenience).

PROPOSITION B.1. *Let $m \geqslant 2$ be an integer and let $f\colon \mathbf{P}^1 \to \mathbf{P}^1, x \mapsto x^m$ be the power map over an algebraically closed field K of characteristic $p > 2$. If p divides m, set $\beta := 0$. Otherwise, let $\beta := (|m^s - 1|_p - 1)/s \in \mathbf{Z}[1/p]$, where s is the smallest positive integer for which $|m^s - 1|_p < 1$ (i.e. the order of m in $\mathbf{F}_p^\times$). Then*

$$\zeta^*_f(z) = \zeta^*_{\mathrm{pt}}(mz)\zeta^*_{\mathrm{pt}}(z)\left(\frac{\zeta^*_{\mathrm{pt}}((mz)^s)}{\zeta^*_{\mathrm{pt}}(z^s)}\right)^{\beta}.$$

PROOF. The iterate f^n has as its fixed points $\infty, 0$, and the distinct solutions to $x^{m^n-1} = 1$ in $\overline{\mathbf{F}}_p$. Hence $f_n = 2 + (m^n - 1)\cdot|m^n - 1|_p$.

If $p|m$, we have $f_n = m^n + 1$, and the result follows. Now assume that m is coprime to p. We then have (see e.g. [**9**, Lemma 6.1])

$$|m^n - 1|_p = \begin{cases} 1 & \text{if } s\nmid n, \\ |m^s - 1|_p \cdot |n|_p & \text{if } s|n. \end{cases} \tag{25}$$

Observe that s is a divisor of $p-1$; in particular, s is coprime to p and $\beta \in \mathbf{Z}[1/p]$. If we set $M := |m^s - 1|_p$, the tame zeta function can be computed as follows:

$$\log \zeta_f^*(z)/\zeta_{\mathrm{pt}}^*(z)^2 = \sum_{p\nmid n; s\nmid n} \frac{m^n - 1}{n} z^n + M \sum_{p\nmid n; s|n} \frac{(m^n - 1)}{n} z^n$$

$$= \sum_n \frac{m^n - 1}{n} z^n - \sum_n \frac{m^{pn} - 1}{pn} z^{pn}$$

$$+ (M-1)\left(\sum_n \frac{m^{sn} - 1}{sn} z^{sn} - \sum_n \frac{m^{psn} - 1}{psn} z^{psn}\right)$$

$$= \log\left(\frac{1-z}{1-mz} \cdot \frac{(1-(mz)^p)^{\frac{1}{p}}}{(1-z^p)^{\frac{1}{p}}}\right) + \frac{(M-1)}{s} \log\left(\frac{1-z^s}{1-(mz)^s} \cdot \frac{(1-(mz)^{ps})^{\frac{1}{p}}}{(1-z^{ps})^{\frac{1}{p}}}\right). \square$$

Without showing further details of computations (that go along the lines of those for the power map) we now list several other tame zeta functions.

PROPOSITION B.2. *Suppose that K is an algebraically closed field of characteristic $p > 2$. Let $m \geqslant 2$ be an integer. The normalised Chebyshev polynomial T_m is the unique monic polynomial of degree m with integer coefficients satisfying $T_m(z + z^{-1}) = z^m + z^{-m}$. Consider the* Chebyshev map

$$T_m \colon \mathbf{P}^1 \to \mathbf{P}^1$$

given by the polynomial T_m (denoted by the same symbol).

Let E/K denote an elliptic curve and let $\pi \colon E \to \mathbf{P}^1$ be a covering map of order two. The (standard) Lattès map

$$L_m \colon \mathbf{P}^1 \to \mathbf{P}^1$$

corresponding to π is defined by the property $L_m \circ \pi = \pi \circ [m]$, where $[m]$ is the multiplication-by-m map on E.

If $p\nmid m$, let s denote the multiplicative order of m modulo p. Let $h = 1$ if f is a Chebyshev map or a Lattès map arising from an ordinary elliptic curve, and let $h = 2$ otherwise. Set $\beta := (|m^s - 1|_p^h - 1)/s \in \mathbf{Z}[1/p]$. Then the corresponding tame zeta functions (quotiented by a convenient factor) are given in Table 2. $\square$

PROPOSITION B.3. *Suppose that K is an algebraically closed field of characteristic $p > 0$ and consider an* additive polynomial *in $K[X]$ of the form $a_0X + a_1X^p + \cdots + a_mX^m$ with $m = p^r$ for some integer $r \geqslant 1$. Assume that $a_0 \in \overline{\mathbf{F}}_p^\times$ and $m \geqslant 2$. Consider f as a map*

$$f \colon \mathbf{P}^1 \to \mathbf{P}^1, X \mapsto a_0X + a_1X^p + \cdots + a_mX^m.$$

Let $s \geqslant 1$ be the smallest integer with $f^s(X) = X + aX^{p^t} + \cdots$ for $a \neq 0$ and $t \in \mathbf{Z}_{>0}$. Put $\beta = (p^{-t} - 1)/s$. Then

$$\zeta_f^*(z) = \zeta_{\mathrm{pt}}^*(mz)\zeta_{\mathrm{pt}}^*(z)\zeta_{\mathrm{pt}}^*((mz)^s)^\beta. \qquad \square$$

Condition	$\zeta^*_{T_m}(z)/(\zeta^*_{\rm pt}(mz)\zeta^*_{\rm pt}(z))$	$\zeta^*_{L_m}(z)/(\zeta^*_{\rm pt}(m^2z)\zeta^*_{\rm pt}(z))$
$p\mid m$	1	1
$p\nmid m$ and $2\nmid s$	$\left(\frac{\zeta^*_{\rm pt}((mz)^s)}{\zeta^*_{\rm pt}(z^s)}\right)^{\beta/2}$	$\left(\frac{\zeta^*_{\rm pt}((m^2z)^s)\zeta^*_{\rm pt}(z^s)}{\zeta^*_{\rm pt}((mz)^s)^2}\right)^{\beta/2}$
$p\nmid m$ and $s=2t$	$\left(\frac{\zeta^*_{\rm pt}((mz)^t)\zeta^*_{\rm pt}(z^t)}{\zeta^*_{\rm pt}(z^{2t})}\right)^{\beta}$	$\left(\frac{\zeta^*_{\rm pt}((m^2z)^t)\zeta^*_{\rm pt}(z^t)\zeta^*_{\rm pt}((mz)^t)^2}{\zeta^*_{\rm pt}((mz)^{2t})^2}\right)^{\beta}$

Table 2. Tame zeta functions of some dynamically affine maps on $\mathbf{P}^1$.

In characteristic two and for more general (sub)additive polynomials, similar methods apply, but the computations are more tedious and we have not listed the outcome. We have not carried out an explicit computation for general Lattès maps arising from endomorphisms of elliptic curves that are not given by multiplication by an integer or corresponding to larger (possibly noncommutative) automorphism groups Γ.

Acknowledgments

We thank Martin Bright, Stefano Marseglia and Sławomir Rams for some illuminating discussions.

References

[1] Satoshi Arima, *Commutative group varieties*, J. Math. Soc. Japan **12** (1960), 227–237, DOI 10.2969/jmsj/01230227. MR0136611

[2] Michael Artin and Barry C. Mazur, *On periodic points*, Ann. of Math. (2) **81** (1965), 82–99, DOI 10.2307/1970384. MR0176482

[3] Robert Auffarth and Giancarlo Lucchini Arteche, *Smooth quotients of abelian varieties by finite groups*, preprint, arXiv:1801.00028 (2018), to appear in Annali della Scuola Normale Superiore di Pisa, Classe di Scienze. DOI 10.2422/2036-2145.201812_008.

[4] Michael Baake, Eike Lau, and Vytautas Paskunas, *A note on the dynamical zeta function of general toral endomorphisms*, Monatsh. Math. **161** (2010), no. 1, 33–42, DOI 10.1007/s00605-009-0118-y. MR2670229

[5] Jason P. Bell, Michael Coons, and Eric Rowland, *The rational-transcendental dichotomy of Mahler functions*, J. Integer Seq. **16** (2013), no. 2, Article 13.2.10, 11. MR3032393

[6] Jason Bell, Richard Miles, and Thomas Ward, *Towards a Pólya-Carlson dichotomy for algebraic dynamics*, Indag. Math. (N.S.) **25** (2014), no. 4, 652–668, DOI 10.1016/j.indag.2014.04.005. MR3217030

[7] Rufus Bowen and Oscar E. Lanford III, *Zeta functions of restrictions of the shift transformation*, Global Analysis (Proc. Sympos. Pure Math., Vol. XIV, Berkeley, Calif., 1968), Amer. Math. Soc., Providence, R.I., 1970, pp. 43–49. MR0271401

[8] Andrew Bridy, *Transcendence of the Artin-Mazur zeta function for polynomial maps of* $\mathbb{A}^1(\overline{\mathbb{F}}_p)$, Acta Arith. **156** (2012), no. 3, 293–300, DOI 10.4064/aa156-3-6. MR2999074

[9] Andrew Bridy, *The Artin-Mazur zeta function of a dynamically affine rational map in positive characteristic* (English, with English and French summaries), J. Théor. Nombres Bordeaux **28** (2016), no. 2, 301–324. MR3509712

[10] Jérôme Buzzi, *Some remarks on random zeta functions*, Ergodic Theory Dynam. Systems **22** (2002), no. 4, 1031–1040, DOI 10.1017/S0143385702000524. MR1925634

[11] Jakub Byszewski and Gunther Cornelissen, *Dynamics on abelian varieties in positive characteristic*, Algebra Number Theory **12** (2018), no. 9, 2185–2235, DOI 10.2140/ant.2018.12.2185. With an appendix by Robert Royals and Thomas Ward. MR3894433

[12] Jakub Byszewski, Gunther Cornelissen, and Marc Houben, *Dynamics on algebraic groups in positive characteristic*, forthcoming, 2019.

[13] Claude Chevalley, *Une démonstration d'un théorème sur les groupes algébriques* (French), J. Math. Pures Appl. (9) **39** (1960), 307–317. MR0126447

[14] Brian Conrad, *A modern proof of Chevalley's theorem on algebraic groups*, J. Ramanujan Math. Soc. **17** (2002), no. 1, 1–18. MR1906417

[15] Pierre Deligne, *Séminaire de Géometrie Algébrique du Bois-Marie (SGA $4\frac{1}{2}$): Cohomologie étale*, Lecture Notes in Math., vol. 569, Springer, Berlin, 1977, avec la collaboration de J.F. Boutot, A. Grothendieck, L. Illusie et J.L. Verdier. MR3727438

[16] Michel Demazure and Pierre Gabriel, *Groupes algébriques. Tome I: Géométrie algébrique, généralités, groupes commutatifs* (French), Masson & Cie, Éditeur, Paris; North-Holland Publishing Co., Amsterdam, 1970. Avec un appendice *Corps de classes local* par Michiel Hazewinkel. MR0302656

[17] Max Deuring, *Die Typen der Multiplikatorenringe elliptischer Funktionenkörper* (German), Abh. Math. Sem. Hansischen Univ. **14** (1941), 197–272, DOI 10.1007/BF02940746. MR0005125

[18] Tien-Cuong Dinh and Nessim Sibony, *Dynamics in several complex variables: endomorphisms of projective spaces and polynomial-like mappings*, Holomorphic dynamical systems, Lecture Notes in Math., vol. 1998, Springer, Berlin, 2010, pp. 165–294, DOI 10.1007/978-3-642-13171-4_4. MR2648690

[19] Bernard Dwork, *On the rationality of the zeta function of an algebraic variety*, Amer. J. Math. **82** (1960), 631–648, DOI 10.2307/2372974. MR140494
4, 905–919]. MR1027462

[20] Alexandre È. Erëmenko, *Some functional equations connected with the iteration of rational functions* (Russian), Algebra i Analiz **1** (1989), no. 4, 102–116; English transl., Leningrad Math. J. **1** (1990), no. 4, 905–919. MR1027462

[21] Alexander L. Fel'shtyn, *Dynamical zeta functions, Nielsen theory and Reidemeister torsion*, Mem. Amer. Math. Soc. **147** (2000), no. 699, xii+146. MR1697460

[22] Leopold Flatto, Jeffrey C. Lagarias, and Bjorn Poonen, *The zeta function of the beta transformation*, Ergodic Theory Dynam. Systems **14** (1994), no. 2, 237–266, DOI 10.1017/S0143385700007860. MR1279470

[23] Giovanni Gallavotti, *Funzioni zeta ed insiemi basilari* (Italian, with English summary), Atti. Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. (8) **61** (1976), no. 5, 309–317 (1977). MR0467842

[24] Mark Giesbrecht and Myung Sub Kim, *Computing the Hermite form of a matrix of Ore polynomials*, J. Algebra **376** (2013), 341–362, DOI 10.1016/j.jalgebra.2012.11.033. MR3003730

[25] Nathan Grieve, *Reduced norms and the Riemann-Roch theorem for Abelian varieties*, New York J. Math. **23** (2017), 1087–1110. MR3711271

[26] Alexander Grothendieck, *Formule de Lefschetz et rationalité des fonctions L* (French), Séminaire Bourbaki, Vol. 9, Soc. Math. France, Paris, 1995, pp. Exp. No. 279, 41–55. MR1608788

[27] Aimo Hinkkanen, *Zeta functions of rational functions are rational*, Ann. Acad. Sci. Fenn. Ser. A I Math. **19** (1994), no. 1, 3–10. MR1246882

[28] Daniel Huybrechts, *Lectures on K3 surfaces*, Cambridge Studies in Advanced Mathematics, vol. 158, Cambridge University Press, Cambridge, 2016. MR3586372

[29] Jamie Juul, Pär Kurlberg, Kalyani Madhu, and Tom J. Tucker, *Wreath products and proportions of periodic points*, Int. Math. Res. Not. IMRN **13** (2016), 3944–3969, DOI 10.1093/imrn/rnv273. MR3544625

[30] Vadim Yu. Kaloshin, *Generic diffeomorphisms with superexponential growth of number of periodic orbits*, Comm. Math. Phys. **211** (2000), no. 1, 253–271, DOI 10.1007/s002200050811. MR1757015

[31] Douglas A. Lind, *A zeta function for $\mathbf{Z}^d$-actions*, Ergodic theory of $\mathbf{Z}^d$ actions (Warwick, 1993), London Math. Soc. Lecture Note Ser., vol. 228, Cambridge Univ. Press, Cambridge, 1996, pp. 433–450, DOI 10.1017/CBO9780511662812.019. MR1411232

[32] Anthony Manning, *Axiom* A *diffeomorphisms have rational zeta functions*, Bull. London Math. Soc. **3** (1971), 215–220, DOI 10.1112/blms/3.2.215. MR0288786
[33] James S. Milne, *Abelian varieties (v2.00)*, 2008, Available at www.jmilne.org/math, 172 pp.
[34] James S. Milne, *Algebraic groups*, Cambridge Studies in Advanced Mathematics, vol. 170, Cambridge University Press, Cambridge, 2017. The theory of group schemes of finite type over a field. MR3729270
[35] Poul G. Hjorth and Carsten Lunde Petersen (eds.), *Dynamics on the Riemann sphere*, European Mathematical Society (EMS), Zürich, 2006. A Bodil Branner Festschrift; Including papers from the Symposium on Holomorphic Dynamics held in Holbæk, June 2003. MR2348951
[36] Violetta B. Pilyugina and A. L. Fel′shtyn, *The Nielsen zeta function* (Russian), Funktsional. Anal. i Prilozhen. **19** (1985), no. 4, 61–67, 96. MR820085
[37] Mark Pollicott, *A complex Ruelle-Perron-Frobenius theorem and two counterexamples*, Ergodic Theory Dynam. Systems **4** (1984), no. 1, 135–146, DOI 10.1017/S0143385700002327. MR758899
[38] Claudio Qureshi and Daniel Panario, *The graph structure of Chebyshev polynomials over finite fields and applications*, Des. Codes Cryptogr. **87** (2019), no. 2-3, 393–416, DOI 10.1007/s10623-018-0545-7. MR3911212
[39] Bernard Randé, *Mahler functional equations and applications to p-regular sequences*, Thesis, Université Bordeaux 1, 1992, Available at HAL https://tel.archives-ouvertes.fr/tel-01183330.
[40] David Ruelle, *Flots qui ne mélangent pas exponentiellement* (French, with English summary), C. R. Acad. Sci. Paris Sér. I Math. **296** (1983), no. 4, 191–193, DOI 10.1142/9789812833709_0024. MR692974
[41] Sanford L. Segal, *Nine introductions in complex analysis*, Revised edition, North-Holland Mathematics Studies, vol. 208, Elsevier Science B.V., Amsterdam, 2008. MR2376066
[42] Jean-Pierre Serre, *Groupes algébriques et corps de classes* (French), Publications de l'institut de mathématique de l'université de Nancago, VII. Hermann, Paris, 1959. MR0103191
[43] Joseph H. Silverman, *The arithmetic of dynamical systems*, Graduate Texts in Mathematics, vol. 241, Springer, New York, 2007. MR2316407
[44] Joseph H. Silverman, *The arithmetic of elliptic curves*, 2nd ed., Graduate Texts in Mathematics, vol. 106, Springer, Dordrecht, 2009. MR2514094
[45] Stephen Smale, *Differentiable dynamical systems*, Bull. Amer. Math. Soc. **73** (1967), 747–817, DOI 10.1090/S0002-9904-1967-11798-1. MR228014
[46] Chris Smyth, *Seventy years of Salem numbers*, Bull. Lond. Math. Soc. **47** (2015), no. 3, 379–395, DOI 10.1112/blms/bdv027. MR3354434
[47] Tonny A. Springer, *Linear algebraic groups*, 2nd ed., Modern Birkhäuser Classics, Birkhäuser Boston, Inc., Boston, MA, 2009. MR2458469
[48] Richard P. Stanley, *Differentiably finite power series*, European J. Combin. **1** (1980), no. 2, 175–188, DOI 10.1016/S0195-6698(80)80051-5. MR587530
[49] Richard P. Stanley, *Enumerative combinatorics. Vol. 2*, Cambridge Studies in Advanced Mathematics, vol. 62, Cambridge University Press, Cambridge, 1999. With a foreword by Gian-Carlo Rota and appendix 1 by Sergey Fomin. MR1676282
[50] Troy Vasiga and Jeffrey Shallit, *On the iteration of certain quadratic maps over* $\mathrm{GF}(p)$, Discrete Math. **277** (2004), no. 1-3, 219–240, DOI 10.1016/S0012-365X(03)00158-4. MR2033734
[51] Wolfram Research, Inc., *Mathematica, Version 11.3*, Champaign, IL, 2018.

Wydział Matematyki i Informatyki Uniwersytetu Jagiellońskiego, ul. S. Łojasiewicza 6, 30-348 Kraków, Polska

Email address: jakub.byszewski@uj.edu.pl

Mathematisch Instituut, Universiteit Utrecht, Postbus 80.010, 3508 TA Utrecht, Nederland

Email address: g.cornelissen@uu.nl

Mathematisch Instituut, Universiteit Utrecht, Postbus 80.010, 3508 TA Utrecht, Nederland

Current address: Departement Wiskunde, KU Leuven, Celestijnenlaan 200b—box 2400, 3001 Leuven, België

Email address: marc.houben@kuleuven.be

Mathematisch Instituut, Universiteit Utrecht, Postbus 80.010, 3508 TA Utrecht, Nederland
Email address: loisvandermeijden@gmail.com

Contemporary Mathematics
Volume **744**, 2020
https://doi.org/10.1090/conm/744/14976

Special α-limit sets

Sergiĭ Kolyada†, Michał Misiurewicz, and L'ubomír Snoha

Abstract. We investigate the notion of the special α-limit set of a point. For a continuous selfmap of a compact metric space, it is defined as the union of the sets of accumulation points over all backward branches of the map. The main question is whether a special α-limit set has to be closed. We show that it is not the case in general. It is unknown even whether a special α-limit set has to be Borel or at least analytic (it is in general an uncountable union of closed sets). However, we answer this question affirmatively for interval maps for which the set of all periodic points is closed. We also give examples showing how those sets may look like and we provide some conjectures and a problem.

1. Introduction

Let (X, f) be a dynamical system given by a compact metric space X and a continuous map $f : X \to X$.

The ω-limit set of $x \in X$, $\omega(x)$, is the set of points of accumulation of the sequence $(f^n(x))_{n=0}^{\infty}$. If f is a homeomorphism, one can define the α-limit set of x, $\alpha(x)$, as the ω-limit set of x for f^{-1}. However, if f is not a homeomorphism, this simple way does not work.

The standard solution is to define the *α-limit set* $\alpha(x)$ of x as the set of limits of all convergent sequences $(x_{n_i})_{i=0}^{\infty}$ such that $f^{n_i}(x_{n_i}) = x$ and $\lim_{i\to\infty} n_i = \infty$. However, one can also think of different solutions.

A *backward branch* of x is a sequence $(x_n)_{n=0}^{\infty}$ such that $x_0 = x$ and $f(x_n) = x_{n-1}$. The *α-limit set of the backward branch* $(x_n)_{n=0}^{\infty}$ is the set of points of accumulation of this sequence.

In 1992, M. Hero [**6**] explored still different path. He defined the *special α-limit set* $s\alpha(x)$ of x as the union of the α-limit sets over all backward branches of x. Clearly, always $s\alpha(x) \subset \alpha(x)$. If we want to specify that we are using the map f, we will write $s\alpha(x, f)$ instead of $s\alpha(x)$ (and similarly $\omega(x, f)$ and $\alpha(x, f)$ instead of $\omega(x)$ and $\alpha(x)$, respectively). We denote $\mathrm{SA}(f) = \bigcup_{x\in X} s\alpha(x, f)$.

To formulate some known facts on special α-limit sets, we recall some notions and fix notations. By $\mathrm{Orb}(x)$ or $\mathrm{Orb}(x, f)$ we denote the orbit of the point x, i.e.

2010 *Mathematics Subject Classification.* Primary 37B20, 37E05; Secondary 54H20.

Key words and phrases. Special α-limit set, interval map.

Research of the second author was partially supported by grant number 426602 from the Simons Foundation. The third author was supported by the Slovak Research and Development Agency under the contract No. APVV-15-0439.

†Deceased May 16, 2018.

the set $\{x, f(x), f^2(x), \dots\}$. By $\mathrm{Fix}(f)$, $\mathrm{Per}(f)$, $\mathrm{Rec}(f)$ and $\Omega(f)$ we denote, respectively, the set of fixed points, periodic points, recurrent points and nonwandering points of f. Recall that x is a *recurrent point* if $x \in \omega(x)$ and it is a *nonwandering point* if for every neighbourhood U of x there is $n \geq 1$ such that $f^n(U) \cap U \neq \emptyset$.

By $\Lambda^\infty = \Lambda^\infty(f)$ we denote the *attracting center* defined as follows. First, for a subset Y of X, put $\Lambda(Y, f) = \bigcup_{x \in Y} \omega(x, f)$. Then $\Lambda^\infty(f) = \bigcap_{n=1}^\infty \Lambda^n(f)$, where $\Lambda^1(f) = \Lambda(X, f)$ and $\Lambda^n(f) = \Lambda(\Lambda^{n-1}(f))$ for every $n > 1$. Further denote $\Gamma(f) = \bigcup_{x \in X} \gamma(x, f)$ where $\gamma(x, f) = \omega(x, f) \cap \alpha(x, f)$. For *interval* maps, by [**13**], we have

$$\mathrm{Per}(f) \subset \mathrm{Rec}(f) \subset \Gamma(f) = \Lambda^2(f) = \Lambda^\infty(f) \subset \overline{\mathrm{Per}(f)} \subset \Lambda^1(f) \subset \Omega(f) \tag{1.1}$$

and, by [**5**], $x \in \Omega(f)$ if and only if $x \in \alpha(x)$.

For a continuous selfmap f of the *interval* $I = [0, 1]$, Hero's main results are the following ones.

(H1) The following are equivalent:
  (1) $x \in s\alpha(y)$ for some y,
  (2) $x \in s\alpha(x)$,
  (3) $x \in \Lambda^\infty$.

(H2) $\mathrm{SA}(f) \setminus \mathrm{Rec}(f) \neq \emptyset$ if and only if f has a periodic point with period not a power of two.

To get (H1) and (H2), Hero proved, among others, the following facts for interval maps.

(H3) If $x \in \mathrm{Rec}(f)$ then $x \in s\alpha(x)$.

(H4) If $x \in \mathrm{SA}(f)$ then $x \in s\alpha(x)$ and $x \in \Gamma(f)$.

(H5) If $x \in \Gamma(f)$ then $x \in s\alpha(x)$.

Combining (H3–H5) with (1.1), we get for interval maps

$$\mathrm{Rec}(f) \subset \mathrm{SA}(f) = \Gamma(f) = \Lambda^2(f) = \Lambda^\infty(f) \subset \Lambda^1(f).$$

Notice that the inclusion $\mathrm{SA}(f) \subset \Lambda^1(f)$ can be reformulated as follows:

(H6) Every special α-limit point is an ω-limit point.[1]

Another interesting result is the following.

(H7) If $y \in s\alpha(x)$ then the orbit closure $\overline{\mathrm{Orb}(y)} = \mathrm{Orb}(y) \cup \omega(y) \subset s\alpha(x)$.

Hero proved (H7) for interval maps but one can see that the proof works in general.

The special α-limit sets have been studied also on graphs and dendrites, see [**10–12**]. Let us also mention that the α-limit sets of the backward branches $(x_n)_{n=0}^\infty$ of a map have been studied in [**3**].

To study special α-limit sets is more complicated than to study ω-limit sets or α-limit sets. While it is clear that the ω-limit sets, α-limit sets and α-limit sets of backward branches are always closed, the situation with the special α-limit sets is unclear. In general, those sets are uncountable unions of closed sets, so a priori their topology may be very complicated. Are those sets closed? If not, are they Borel or at least analytic? And are there any other constraints on them? Those questions can be asked in a general case, but as always, the special case when $X = I = [0, 1]$

[1] This can be alternatively proved as a trivial consequence of a Sharkovsky's result, see [**9**], saying that a point $c \in [0, 1]$ lies in $\Lambda^1(f)$ if and only if every open interval containing c contains at least three points of some trajectory. Recall another trivial consequence: the set $\Lambda^1(f)$ is closed.

is the closed unit interval promises more results than the general one. Thus, in this paper we will concentrate mainly on this special case.

Our partial answers to the main question whether the set $s\alpha(x)$ is necessarily closed, are as follows.

- In the general case $s\alpha(x)$ does not have to be closed, see Example 2.1.
- For interval maps with closed set of periodic points, $s\alpha(x)$ has to be closed, see Theorem 3.3.
- For transitive interval maps $s\alpha(x)$ has to be closed, see Corollary 3.11.

Our conjecture is as follows.

CONJECTURE 1.1. *For all continuous maps* $[0,1] \to [0,1]$ *all sets* $s\alpha(x)$ *are closed.*

Further, we show that

- not all closed subsets of $[0,1]$ are special α-limit sets for interval maps, see Proposition 3.6, see also Proposition 3.7.

Therefore it is natural to state the following problem.

PROBLEM 1.2. *Characterize all subsets/closed subsets* A *of* $[0,1]$ *for which there exists a continuous map* $[0,1] \to [0,1]$ *and a point* $x \in [0,1]$ *such that* $s\alpha(x) = A$.

For completeness recall that the characterization of ω-limit sets for continuous interval maps is nontrivial but known, see [**1**]: A subset of $[0,1]$ is an ω-limit set for some continuous map $[0,1] \to [0,1]$ if and only if it is nonempty, closed and either nowhere dense or a union of finitely many nondegenerate closed intervals.

Characterization of α-limit sets for continuous interval maps is trivial (and apparently not mentioned in literature): A subset $A \subset [0,1]$ is an α-limit set for some continuous map $[0,1] \to [0,1]$ if and only if it is closed (possibly empty). One implication is trivial. To prove the converse implication, fix a closed set $A \subset [0,1]$. We need to find a continuous map f and a point x such that $\alpha(x) = A$. Examples with $A = \emptyset$ and $A = [0,1]$ are easy (a non-surjective map, cf. Proposition 2.3, and for instance the full tent map, respectively). Otherwise, let J be a connected component of the open set $[0,1] \setminus A$. At least one of the endpoints of J belongs to A; denote it by x. Let f be a continuous map with $f(A) = \{x\}$ and $f([0,1] \setminus A) \subset J$. Then $\bigcup_{n=1}^{\infty} f^{-n}(x) = A$ and $\alpha(x) = A$.

2. General case

We start by providing an example where a special α-limit set is not closed. It will always be clear from the context whether $(0,1)$ means an open interval or the point of the plane with coordinates 0 and 1.

EXAMPLE 2.1. First for every $a \in (0,1)$ we define the map $\varphi_a : [0,1] \to [0,1]$ by setting $\varphi_a(x) = \min(x/a, 1)$ (see Figure 1). Clearly, the maps φ_a converge uniformly to the identity as a goes to 1.

Now our space $X \subset \mathbb{R}^2$ is the union of straight line segments I_a joining the point $(0,1)$ with $(a,0)$ over all $a \in \{1 - 1/n : n = 2,3,4,\dots\} \cup \{1\}$. We define the map $f : X \to X$ by defining f on I_1 as the identity and on I_a for $a < 1$ as $\pi_a^{-1} \circ \varphi_a \circ \pi_a$, where $\pi_a : I_a \to [0,1]$ is the projection to the second coordinate. Clearly, f is continuous.

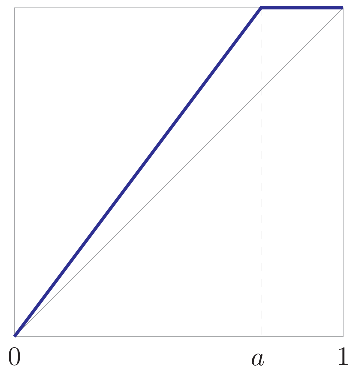


FIGURE 1. The map φ_a.

For every $a \in (0,1)$ we have $s\alpha(1, \varphi_a) = \{0, 1\}$, while for the identity $s\alpha(1, \mathrm{Id}) = \{1\}$. Thus, we get $s\alpha((0,1), f) = \{(0,1)\} \cup \{(1-1/n, 0) : n = 2, 3, 4, \dots\}$. This set is not closed.

Note that we can modify the above example by replacing the set $\{1-1/n : n = 2, 3, 4, \dots\}$ by the interval $[1/2, 1)$, and then our phase space will be a triangle.

Now we prove some basic facts that hold in the general case. We work here with a dynamical system (X, f) where X is a compact metric space and $f\colon X \to X$ is continuous.

In general, $s\alpha(x)$ (as well as $\alpha(x)$) may be empty, because for instance it may happen that x has no preimages. Let us investigate closer this problem. For this we need a simple topological lemma.

LEMMA 2.2. *If* $Y = \bigcap_{n=0}^{\infty} f^n(X)$ *then* $f(Y) = Y$.

PROOF. Clearly, $f(Y) \subset Y$. Take a point $y \in Y$. Then for every n we have $y \in f^{n+1}(X)$, so there is $x_n \in f^n(X)$ such that $f(x_n) = y$. From the sequence (x_n) we can choose a subsequence convergent to some $x \in X$. For each m almost all points x_n belong to $f^m(X)$, and therefore $x \in f^m(X)$. Thus, $x \in Y$. By continuity, $f(x) = y$. This proves that $Y \subset f(Y)$. □

Without compactness of X the lemma would be false. Take the discrete space consisting of z, y and $x_{n,m}$ where n, m are integers, $1 \le n \le m$. Set $f(z) = f(y) = z$, $f(x_{1,m}) = y$ and $f(x_{n,m}) = x_{n-1,m}$ if $n > 1$. Then $\{y, z\} = Y \neq f(Y) = \{z\}$.

The following proposition gives the answer to the question when $s\alpha(x) \neq \emptyset$.

PROPOSITION 2.3. *The following conditions are equivalent:*

(1) $s\alpha(x) \neq \emptyset$,
(2) $\alpha(x) \neq \emptyset$,
(3) $x \in \bigcap_{n=0}^{\infty} f^n(X)$.

In particular, if f is a surjection then $s\alpha(x)$ is nonempty for every $x \in X$.

PROOF. Since $s\alpha(x) \subset \alpha(x)$, (1) implies (2). By the definition, if (2) holds then $f^{-n}(x) \neq \emptyset$ for all n, and thus (3) holds. Hence, it remains to prove that (3) implies (1).

By Lemma 2.2, if $x \in \bigcap_{n=0}^{\infty} f^n(X)$ then x has an infinite backward branch. From this branch we can choose a convergent subsequence, so $s\alpha(x) \neq \emptyset$. □

Recall that $f : X \to X$ is *minimal* if every orbit is dense and is *topologically exact* if for every nonempty open set $U \subset X$ there is a positive integer n such that $f^n(U) = X$. Further, a set $E \subset X$ is f-invariant if $f(E) \subset E$. By $B(x, \delta)$ denote the ball with radius δ, centered at x.

PROPOSITION 2.4. *Let $E \subset X$ be an f-invariant set such that $\bigcup_{n=1}^{\infty} f^n(U) \supset X \setminus E$ for every nonempty open set $U \subset X$. Then*

$$s\alpha(x) = X \quad \text{for every } x \in X \setminus E.$$

Thus, if $E = \emptyset$ then $s\alpha(x) = X$ for every $x \in X$. In particular, this is true if f is topologically exact or minimal.

PROOF. Fix $x \in X \setminus E$ and $y \in X$. We prove that $y \in s\alpha(x)$.

Due to the assumption, in the ball $B(y, 1)$ there is a point y_1 whose f^{k_1}-image (for some $k_1 \geq 1$) is x. Since E is f-invariant and $x \in X \setminus E$, also $y_1 \in X \setminus E$. By the same argument as above, there is a point $y_2 \in B(y, 1/2) \setminus E$ with $f^{k_2}(y_2) = y_1$ (for some $k_2 \geq 1$). Continuing this way, we find a sequence of points $(y_n)_{n=1}^{\infty}$ in $X \setminus E$ and a sequence of positive integers $(k_n)_{n=1}^{\infty}$ such that $y_n \to y$ and $f^{k_n}(y_n) = y_{n-1}$ for $n \geq 2$ and $f^{k_1}(y_1) = x$. Hence $y \in s\alpha(x)$. □

Note that if the set E in this proposition is not equal to X then it has *empty interior*. Otherwise it contains a nonempty open set U for which then trivially $\bigcup_{n=1}^{\infty} f^n(U) \subset E$. Since $E \neq X$, this contradicts the assumption that $\bigcup_{n=1}^{\infty} f^n(U) \supset X \setminus E$. The proposition thus shows that, under those assumptions, for almost every (in topological sense) point x, the set $s\alpha(x)$ equals the whole space X.

Let us now study what happens if we replace a point with its image or a preimage.

LEMMA 2.5. *If $y \in s\alpha(x)$ then*

(1) $\overline{\mathrm{Orb}(y)} \subset s\alpha(x)$*; in particular* $f(y) \in s\alpha(x)$,
(2) $y \in s\alpha(f(x))$ *and* $f(y) \in s\alpha(f(x))$.

PROOF. (1) This follows from (H7) and our comment after it.

(2) The first property follows from the definition and the second one is a consequence of it and of (1). □

LEMMA 2.6. *If $y \in s\alpha(x)$ then $f^{-1}(y) \cap s\alpha(x) \neq \emptyset$.*

PROOF. If $y \in s\alpha(x)$ then there is a sequence of points $(y_n)_{n=1}^{\infty}$ and a sequence of positive integers $(k_n)_{n=1}^{\infty}$ such that $y_n \to y$ and $f^{k_n}(y_n) = y_{n-1}$ for $n \geq 2$ and $f^{k_1}(y_1) = x$. Set $z_n = f^{k_n - 1}(y_n)$. By passing to a subsequence if necessary, we may assume that $z_n \to z$. Then $f(z) = \lim_{n\to\infty} f(z_n) = \lim_{n\to\infty} f^{k_n}(y_n) = \lim_{n\to\infty} y_{n-1} = y$. Hence $z \in f^{-1}(y)$. Since $f^{k_n}(z_{n+1}) = z_n$ for $n \geq 1$ and $f(z_1) = x$, we also have $z \in s\alpha(x)$. □

From Lemmas 2.5 and 2.6 we get immediately the following corollary.

COROLLARY 2.7. *We have $f(s\alpha(x)) = s\alpha(x)$. In particular, if $s\alpha(x)$ is finite then it is a union of periodic orbits.*

The following lemma will be useful when constructing examples. By $\mathrm{Int}\, C$ we denote the interior of C.

LEMMA 2.8. *Suppose that there is an f-invariant (not necessarily closed or open) set C and a point $y \in X$ with the following two properties:*

(a) *y has positive distance from C,*
(b) *y is pre-trapped by C, meaning that there exists a positive integer t such that $f^t(y) \in \operatorname{Int} C$.*

Then $y \notin \mathrm{SA}(f) := \bigcup_{x \in X} s\alpha(x)$.

PROOF. Suppose, on the contrary, that $y \in s\alpha(x)$ for some x. Then there is a sequence of points $(y_n)_{n=1}^{\infty}$ and a sequence of positive integers $(k_n)_{n=1}^{\infty}$ such that $y_n \to y$, $f^{k_n}(y_n) = y_{n-1}$ for $n \geq 2$ and $f^{k_1}(y_1) = x$. Since f is continuous, by (a) and (b) there is a neighbourhood U of y such that $U \cap C = \emptyset$ and $f^t(U) \subset C$. Then, since $f(C) \subset C$, we get $f^{\ell}(U) \subset C$ for all $\ell \geq t$. Since $y_n \to y$, there is n_0 such that for all $n \geq n_0$ we have $y_n \in U$. Thus,

$$n \geq n_0 \text{ and } \ell \geq t \Rightarrow f^{\ell}(y_n) \in C. \tag{2.1}$$

Consider $t+1$ points $y_{n_0}, y_{n_0+1}, \dots, y_{n_0+t} \in U$. For $L := k_{n_0+1} + k_{n_0+2} + \dots + k_{n_0+t}$ we have

$$f^L(y_{n_0+t}) = y_{n_0} \in U. \tag{2.2}$$

However, $n_0+t \geq n_0$ and for the integer L, which is the sum of t positive integers, we have $L \geq t$. Therefore we can use (2.1) to get $f^L(y_{n_0+t}) \in C$. This contradicts (2.2). □

Now we investigate how the special α-limit sets behave when we consider the iterations of the map.

PROPOSITION 2.9. *Let n be a positive integer. The special α-limit sets have the following properties:*

(1) $s\alpha(x, f^n) \subset s\alpha(x, f)$,
(2) $s\alpha(x, f) \subset s\alpha(f^n(x), f)$,
(3) $s\alpha(x, f) = \bigcup_{i=0}^{n-1} \bigcup_{f^i(y)=x} s\alpha(y, f^n)$,
(4) $s\alpha(x, f) \subset \bigcup_{j=0}^{n-1} s\alpha(f^j(x), f^n)$.
(5) $\mathrm{SA}(f^n) = \mathrm{SA}(f)$.

PROOF. Property (1) follows immediately from the definition, while (2) follows from Lemma 2.5. To prove (3), observe that a backward branch of x for f decomposes into n backward branches of preimages of x for f^n. Property (4) follows from (3) and (2). Finally, one inclusion in (5) follows from (1) and the other one from (3) or (4). □

3. Interval maps

Here we prove the main results about the special α-limit sets for interval maps. Thus, $I = [0, 1]$ will be the closed unit interval, and $f : I \to I$ a continuous map.

LEMMA 3.1. *Let f be an interval map. If $y \in \alpha(x)$ and $f(y) = y$, then $y \in s\alpha(x)$.*

PROOF. If $y = x$, then $f(x) = x$, so $y = x \in s\alpha(x)$. Thus, we may assume that $y < x$.

By the assumption, there exists a sequence $x_n \to y$ such that $f^{k_n}(x_n) = x$ for some $k_n \to \infty$. By choosing a subsequence, we may assume that the sequence (x_n) is monotone and $x_n < x$ for all n.

Assume first that the sequence (x_n) is decreasing. By induction we will find a sequence (y_n) such that $f^m(y_1) = x$ for some m, and $f^{k_n}(y_n) = y_{n-1}$, $y < y_n < x_n$, for all $n \geq 2$. We start by taking $y_1 = x_1$ and $m = k_1$. Now, if y_{n-1} is already chosen, we note that $y_{n-1} \in (y, x)$, $f^{k_n}(y) = y$, and $f^{k_n}(x_n) = x$, so there exists $y_n \in (y, x_n)$ such that $f^{k_n}(y_n) = y_{n-1}$. This completes the induction step. Clearly, $y_n \to y$, so by the definition, $y \in s\alpha(x)$.

Now we assume that the sequence (x_n) is increasing. Then $x_n < y < x$ and $f^{k_n}(x_n) = x$, so there exists $0 \leq m_n < k_n$ such that $f^{m_n}(x_n) < y$ and $f^{m_n+1}(x_n) > y$. We will distinguish two cases.

The first case is that $\sup_n f^{m_n}(x_n) < y$. Denote this supremum by a. Since $x_n \to y$, we may assume that $m_n > 0$ and $x_n > a$ for all n. Then we repeat the proof from the case of (x_n) decreasing. Namely, by induction we find a sequence (y_n) such that $f^m(y_1) = x$, and $f^{m_n}(y_n) = y_{n-1}$, $x_n < y_n < y$, for all $n \geq 2$. This time in the induction step we have $y_{n-1} \in (a, y)$, $f^{m_n}(y) = y$, and $f^{m_n}(x_n) \leq a$. Again we get $y \in s\alpha(x)$.

The second case is that $\sup_n f^{m_n}(x_n) = y$. Then, by passing to a subsequence, we may assume that $f^{m_n}(x_n) \to y$. By continuity and since $f(y) = y$, we get $f^{m_n+1}(x_n) \to y$. By taking a subsequence again, we may assume that the sequence $(f^{m_n+1}(x_n))$ is decreasing. Notice that $f^{k_n-m_n-1}(f^{m_n+1}(x_n)) = x$. Therefore, if $k_n - m_n - 1 \to \infty$, we can use the sequence $(f^{m_n+1}(x_n))$ instead of the sequence (x_n), and we know already that in this case $y \in s\alpha(x)$. Thus, it remains to show that the sequence $(k_n - m_n - 1)$ of nonnegative integers is unbounded (then a subsequence tends to infinity and we may use the mentioned argument). Suppose, on the contrary, that the sequence $(k_n - m_n - 1)$ is bounded. By passing to a subsequence we may assume that it is constant, i.e. there is $N \geq 0$ such that $k_n - m_n - 1 = N$, $n \in \mathbb{N}$. Since $f^{m_n+1}(x_n) \to y$, by continuity we get $f^N(y) = x$, which is impossible since y is a fixed point of f. □

THEOREM 3.2. *Let f be an interval map. If a point $y \in \alpha(x)$ is periodic, then $y \in s\alpha(x)$.*

PROOF. Let $y \in \alpha(x, f)$ be a periodic point of f of period p. Then there exists a sequence of points (y_k) convergent to y, such that $f^{m_k}(y_k) = x$ for some sequence (m_k) going to infinity. There is $j \in \{0, 1, \ldots, p-1\}$ such that infinitely many numbers m_k are congruent to j modulo p. By passing to a subsequence if necessary, we may assume that for every k there is a positive r_k with $m_k = pr_k + j$. Put $z_k = f^j(y_k)$. By continuity, $z_k \to f^j(y)$. Moreover, $r_k \to \infty$ and for $g = f^p$ we have $g^{r_k}(z_k) = x$. Hence, $f^j(y) \in \alpha(x, g)$. Since $f^j(y)$ is a fixed point of g, Lemma 3.1 gives us $f^j(y) \in s\alpha(x, g) = s\alpha(x, f^p)$. Then, by Proposition 2.9(1), $f^j(y) \in s\alpha(x, f)$ and, by Lemma 2.5(1), $y = f^{p-j}(f^j(y)) \in s\alpha(x, f)$. □

Once we know that $\alpha(x) \cap \mathrm{Per}(f) \subset s\alpha(x)$, we can specify a class of interval maps for which all special α-limit sets are closed.

THEOREM 3.3. *For an interval map f, if the set of all periodic points is closed, then $s\alpha(x) = \alpha(x) \cap \mathrm{Per}(f)$ for every x. In particular, all special α-limit sets are closed and $\mathrm{SA}(f) = \mathrm{Per}(f)$.*

Proof. Let f be an interval map for which the set $\mathrm{Per}(f)$ is closed. Then, by a theorem of Sharkovsky (see [**8**]), the ω-limit set of every point is finite and so it is a periodic orbit. Thus, by (H6), all special α-limit points are contained in $\mathrm{Per}(f)$. Hence, for any point x we have $s\alpha(x) \subset \alpha(x) \cap \mathrm{Per}(f)$. Since Theorem 3.2 gives the converse inclusion, $s\alpha(x) = \alpha(x) \cap \mathrm{Per}(f)$. Since both $\alpha(x)$ and $\mathrm{Per}(f)$ are closed, the set $s\alpha(x)$ is also closed. Since every $s\alpha(x) \subset \mathrm{Per}(f)$ and also $\mathrm{Per}(f) \subset \mathrm{SA}(f)$, we have $\mathrm{SA}(f) = \mathrm{Per}(f)$. □

Notice that the set $\alpha(x)$ may be larger than $s\alpha(x)$ even under the assumptions of Theorem 3.3, see Examples 4.3 and 4.4.

An interval map is called *of type* n if it has a periodic orbit of period n but no periodic orbits of periods preceding n in the Sharkovsky ordering. Additionally, it is *of type* 2^∞ if it has periodic orbits of period 2^n for all n and of no other periods. If f is of type 2^n for some finite n, then $\mathrm{Per}(f)$ consists of all fixed points of f^{2^n}, so it is closed. Therefore, we get a corollary to Theorem 3.3.

Corollary 3.4. *For an interval map of type 2^n for some finite n, all special α-limit sets are closed.*

This is still true for interval maps of type 2^∞ with closed set of periodic points. However, there are well known examples of interval maps of type 2^∞ whose set of periodic points is not closed (see, e.g., [**8**]). Recall also that for the maps having also periodic points whose period is not a power of two, the set of periodic points is never closed.

Let us now continue with other properties of the special α-limit sets for interval maps (so we assume below that $f : [0,1] \to [0,1]$ is a continuous map). By an *interval* we always mean a *nondegenerate* interval (still, we sometimes emphasize that it is nondegenerate).

Lemma 3.5. *If f is constant on some open interval J, then J contains at most one point of $s\alpha(x)$; this point is periodic and its orbit contains x. In particular, if $K \subset s\alpha(x)$ is an interval, then $f^n(K)$ is an interval (and not a singleton) for every positive integer n.*

Proof. Assume that f is constant on an open interval J, and $y \in J$ belongs to $s\alpha(x)$. Then there are points $z, w \in J$ such that $f^k(w) = z$ and $f^n(z) = x$ for some $k, n \geq 1$. Thus, $f^k(y) = f^k(z) = f^k(w) = z$ and $f^n(y) = f^n(z) = f^n(w) = x$. Hence $f^{k+n}(y) = x$. We get $x = f^k(f^n(y)) = f^k(x)$, so x is periodic. Moreover, $f^k(z) = z$, so z is also periodic. However, $f^n(z) = x$, so x and z must belong to the same periodic orbit. Since $y \in s\alpha(x)$, the point z can be chosen arbitrarily close to y, and this proves that y belongs to the orbit of x. Thus y is periodic. Clearly, the orbit of x, since it is periodic, can contain at most one point of J.

Now let $K \subset s\alpha(x)$ be an interval. We already know that then f is not constant on K, i.e. $f(K)$ is an interval. By Corollary 2.7 we have $f(K) \subset s\alpha(x)$, hence also $f^2(K)$ is an interval. By induction, $f^n(K)$ is an interval for every n. □

A subinterval K of $[0,1]$ is called *periodic* or *weakly periodic* if there is a positive integer k such that $K, f(K), \dots, f^{k-1}(K)$ are pairwise disjoint and $f^k(K) = K$ or $f^k(K) \subset K$, respectively. In such a case, the set $\bigcup_{i=0}^{k-1} f^i(K)$ is called a *cycle* or a *weak cycle*, respectively, of intervals of period k.

Proposition 3.6. *Assume that* $\mathrm{Int}\, s\alpha(x) \neq \emptyset$. *Then there are only finitely many nondegenerate components of $s\alpha(x)$ and they form a cycle containing x.*

PROOF. Let J be a nondegenerate component of $s\alpha(x)$. Then there are $y, z \in \operatorname{Int}(J)$ such that $f^n(y) = z$ for some $n \geq 1$. Since, by Corollary 2.7, $f(s\alpha(x)) = s\alpha(x)$, we have $f^n(J) \subset J$ and by Lemma 3.5, $f^i(J)$, $i = 1, \dots, n-1$ are nondegenerate intervals. Thus, we get a weak cycle of intervals in $s\alpha(x)$ of period at most n. If we enlarge those intervals to the components of $s\alpha(x)$ containing them, we get a weak cycle of nondegenerate components of $s\alpha(x)$. Since some points of this weak cycle are eventually mapped to x, the point x itself has to belong to one of those components. If there are two different weak cycles of nondegenerate components of $s\alpha(x)$, one of them does not contain x, a contradiction.

Thus there are an interval K and a positive integer k such that the union of all nondegenerate components of $s\alpha(x)$ equals $\bigcup_{i=0}^{k-1} f^i(K)$, where $K, f(K), \dots, f^{k-1}(K)$ are pairwise disjoint intervals and the interval $f^k(K) \subset K$. We claim that this weak cycle is a cycle, i.e. $f^k(K) = K$. Suppose, on the contrary, that $f^k(K)$ is a proper subinterval of K. Then, regardless of whether K is closed or not, $\operatorname{Int}(K \setminus f^k(K)) \neq \emptyset$. Let $L \subset K \setminus f^k(K)$ be an open interval. Then $L \subset s\alpha(x)$ and so, by Lemma 3.5, $f^k(L) \subset f^k(K)$ is a nondegenerate interval. Therefore one can choose a point $y \in L$ with $f^k(y) \in \operatorname{Int} f^k(K)$. Since L is open, y has positive distance from $f^k(K)$. Moreover, $f^k(K) \subset K$ and so the set $f^k(K)$ is f^k-invariant. Hence, by Lemma 2.8, $y \notin \mathrm{SA}(f^k)$. By Proposition 2.9(5), $y \notin \mathrm{SA}(f)$. This contradicts the fact that $y \in L \subset s\alpha(x)$. □

Now we will consider simultaneously two situations; if at least one of them occurs, we will say that x is *of cyclic type*. The first one is when the orbit of x is periodic, the second one when there is a nondegenerate component of $s\alpha(x)$. In the second situation, by Proposition 3.6, those components form a cycle and x belongs to one of them. We will call the union of those components (or the periodic orbit of x) the *cycle of* x.

In both cases we can use the standard techniques of combinatorial dynamics (see, e.g., [**2**]). The closures of the components of the complement of the cycle of x in the convex hull of this cycle are vertices of a *Markov graph*. The graph is directed; there is an arrow from J to K if K is contained in the interval with endpoints $f(a)$ and $f(b)$, where $J = [a, b]$. Then we can use the symbolic dynamics. In particular, for every path in the graph there is a point of the interval with that itinerary.[2] The orbit of this point x goes along the path, that is, $f^n(x)$ belongs to the nth vertex of the path.

Moreover, if the path is finite, of length k, then we can prescribe any point of the kth vertex as $f^k(x)$. Different infinite paths result in different points, except for at most countably many paths. Indeed, if a point y has two different infinite itineraries, then $f^n(y) = x$ for some n, and x is a periodic point. If the period of x is k, then the itinerary of y, starting at $(n+k)$th term or earlier, becomes periodic and goes along the *fundamental loop* of the Markov graph (see [**4**]). For a given n there are only finitely many such itineraries, so totally this can happen only for countably many itineraries.[3]

We can also speak of the *entropy* of the cycle (or of the graph). It is the minimal possible topological entropy of a continuous interval map with this cycle.

[2]The converse is not true. Due to our definition of the Markov graph, some points may be mapped by f in a way which does not correspond to arrows in the graph.

[3]This is the same situation as with the decimal expansions of the numbers from $[0, 1]$; if we remove countably many expansions, then two different expansions give different numbers.

It is equal to the logarithm of the spectral radius of the transition matrix of its Markov graph (see Corollary 4.4.8 of [**2**]). In particular, if the period of the cycle is not a power of 2, then its entropy is positive (see Corollary 4.4.18 of [**2**]).

Proposition 3.7. *Assume that x is of cyclic type and the entropy of its cycle is positive. Then the cardinality of $s\alpha(x)$ minus the cycle of x is continuum.*

Proof. Clearly, this cardinality cannot be larger than continuum. We will show that it is at least continuum.

Since the entropy of the Markov graph is positive, it is easy to see that there is a transitive subgraph G of positive entropy (*transitive* means that there is a path from every vertex to every one). Indeed, if the entropy is positive then there are two distinct loops through some vertex; taking only vertices appearing in those loops produces a transitive subgraph with positive entropy. In this situation, there is a vertex J of G from which there are arrows to at least two vertices. This means that $f(J)$ contains at least two vertices, so it contains some element of our cycle. Thus, there is a point $x_0 \in J$ such that $f^j(x_0) = x$ for some $j \geq 0$.

We will write paths as strings of vertices. Let $C = V_0V_1V_2\dots$ be an infinite path in the graph G. Let us construct by induction a sequence of finite paths C_n as follows. Set $C_0 = J$ (a path of length 0; the length of a path is the number of arrows in it). If the path C_{n-1} is defined, then the path C_n is the concatenation of three paths: the beginning of C of length n, i.e. $V_0V_1\dots V_n$, then a connector $W_1^n \dots W_{s(n)}^n$, and then the path C_{n-1}. The connector is chosen in such a way that C_n is a path in G; this is possible since G is transitive.

Recall that $f^j(x_0) = x$ where $x_0 \in J$. During the induction step we can also choose a point $x_n \in V_0$ such that $f^{n+s(n)+1}(x_n) = x_{n-1}$ (remember that x_{n-1} lies in the first vertex of C_n after the connector, i.e. in J if $n = 1$ and in V_0 if $n \geq 2$). In such a way, the sequence (x_n) is a subsequence of the backward branch of x. From this sequence we can choose a subsequence convergent to some point y. Then $y \in s\alpha(x)$. Since $x_m \in V_0$ for all $m \geq 1$, we have $y \in V_0$. Further, if $n \geq 1$ then for all $m \geq n$ the point $f^n(x_m)$ belongs to V_n. Thus, $f^n(y) \in V_n$. Therefore, the path corresponding to y is C.

Since the entropy of G is positive, the number of infinite paths in G has cardinality continuum. Thus, we get a subset of $s\alpha(x)$ of cardinality continuum, and only finitely many of those points can belong to our cycle of points or intervals (if such a point belongs to an interval of the cycle, it has to be its endpoint). □

From Propositions 3.6 and 3.7 we get immediately two corollaries.

Corollary 3.8. *Assume that x is periodic and $s\alpha(x)$ has empty interior. If the orbit of x has positive entropy (in particular, if its period is not a power of 2) then the cardinality of the set of components of $s\alpha(x)$ is continuum.*

Corollary 3.9. *Assume that $s\alpha(x)$ has nonempty interior and the cycle of the nondegenerate components of $s\alpha(x)$ has positive entropy (in particular, if its period is not a power of 2). Then the cardinality of the set of components of $s\alpha(x)$ that are singletons is continuum.*

Now we are going to study special α-limit sets of transitive interval maps. Recall that if $f\colon [0,1] \to [0,1]$ then the endpoint 0 (resp. 1) is *accessible* if there exists $x \in (0,1)$ and $n \geq 1$ such that $f^n(x) = 0$ (resp. $f^n(x) = 1$). If an endpoint is not accessible, it is called *nonaccessible*. Recall also, see e.g. [**7**, Section 2.1.4], that a transitive interval map is of one of the following two kinds:

- A topologically mixing map. Let $E \subset \{0,1\}$ be the set of those endpoints which are nonaccessible. It can be $E = \emptyset$ or $E = \{0\}$ or $E = \{1\}$ or $E = \{0,1\}$. Always $f(E) = E$ (if $E = \{0,1\}$, then the endpoints are either fixed points or they form a periodic orbit of period 2) and it follows from the definition of an accessible point that also $f^{-1}(E) = E$. Moreover, for every nonempty open set $U \subset [0,1]$, $\bigcup_{n=1}^{\infty} f^n(U) \supset [0,1] \setminus E$.
- A transitive map which is not topologically mixing. Then there is a point $c \in (0,1)$ such that $f([0,c]) = [c,1]$, $f([c,1]) = [0,c]$ and both $f^2|_{[0,c]}$ and $f^2|_{[c,1]}$ are topologically mixing (clearly, then c is the unique fixed point of f). Let E_1 or E_2 be the set of nonaccessible endpoints for $f^2|_{[0,c]}$ or $f^2|_{[c,1]}$, respectively, and let $E = E_1 \cup E_2$. Then $E \subset \{0,c,1\}$ and both $f(E) = E$ and $f^{-1}(E) = E$. Again, for every nonempty open set $U \subset [0,1]$, $\bigcup_{n=1}^{\infty} f^n(U) \supset [0,1] \setminus E$.

PROPOSITION 3.10. *Let f be a transitive interval map. Except of at most three points, for all other x we have $s\alpha(x) = [0,1]$. For each of those (at most three) exceptional points, $s\alpha(x)$ is nonempty and consists of at most two points.*

PROOF. The set E from the above classification of transitive maps has at most three points, is invariant and $\bigcup_{n=1}^{\infty} f^n(U) \supset [0,1] \setminus E$ for every nonempty open set $U \subset [0,1]$. Now use Proposition 2.4 to get that $s\alpha(x) = [0,1]$ for every $x \in [0,1] \setminus E$. Since $f^{-1}(E) = E$ and E is closed, for every $x \in E$ we have $s\alpha(x) \subset E$ (here $s\alpha(x) \neq \emptyset$ since f is surjective). If $E = \{0,c,1\}$, then we cannot have $s\alpha(x) = E$, since then $f^{-1}(\{c\}) = \{c\}$ and $f^{-1}(\{0,1\}) = \{0,1\}$. □

COROLLARY 3.11. *Let f be a transitive interval map. If $s\alpha(x)$ contains more than two points then $s\alpha(x) = [0,1]$. In particular, for every $x \in [0,1]$ the set $s\alpha(x)$ is closed.*

4. Examples

In this section we provide examples of special α-limit sets of various kinds for interval maps. We also formulate some conjectures for interval maps.

EXAMPLE 4.1. While an ω-limit set for a continuous interval map cannot be the disjoint union of a closed interval and a singleton, Figure 2 shows that this is possible for a special α-limit set.

EXAMPLE 4.2. A preimage of a point in $s\alpha(x)$ need not belong to $s\alpha(x)$. In Figure 2 there are points outside $s\alpha(x)$ which are mapped to $s\alpha(x)$. However, see Lemma 2.6.

EXAMPLE 4.3. For maps φ_a from Example 2.1, we have $s\alpha(1) = \{0,1\}$, while $\alpha(1) = [0,1]$ is much larger.

EXAMPLE 4.4. A constant piece in the previous example is not that important, see Figure 3.

EXAMPLE 4.5. Let $f(0) = 0$, $f(1/4) = f(3/4) = 1/2$, $f(1) = 1$ and let f be linear in between, see Figure 4. Then it is a continuous map $[0,1] \to [0,1]$ and $s\alpha(1/2) = \{0,1/2,1\}$, all the three points being fixed for f.

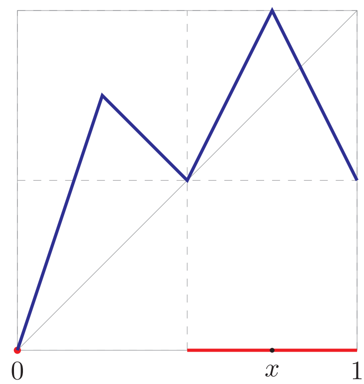


FIGURE 2. Here $s\alpha(x) = \{0\} \cup [1/2, 1]$.

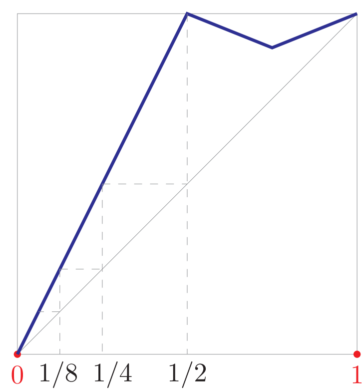


FIGURE 3. Here $s\alpha(1) = \{0, 1\}$ and $\alpha(1) = \{1/2^{n-1} : n \in \mathbb{N}\} \cup \{0\}$.

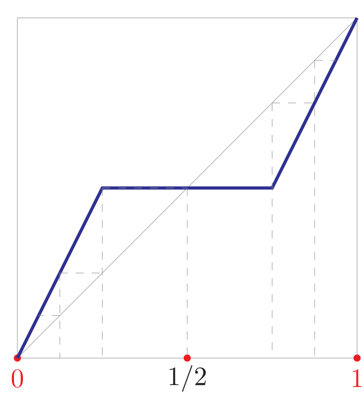


FIGURE 4. Here $s\alpha(1/2) = \{0, 1/2, 1\}$.

EXAMPLE 4.6. For a continuous map $f\colon [0,1] \to [0,1]$, $s\alpha(x)$ can be countable infinite. To get $s\alpha(1) = \{1, 1/2, 1/4, \dots\} \cup \{0\}$, consider the map from Figure 5.

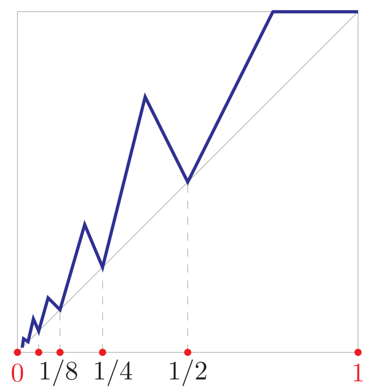


FIGURE 5. Here $s\alpha(1)$ is countable infinite.

EXAMPLE 4.7. We construct a continuous map $f : [0,1] \to [0,1]$ with $s\alpha(1)$ being the middle third Cantor set, see Figure 6. In the construction, the peak over the middle third goes to the very top. In each step, only the peak over the rightmost contiguous interval of that rank goes to the very top. The heights of other peaks tend to zero as the rank tends to infinity and so we get a continuous map. On the other hand, those heights are chosen in such a way that the following property is fulfilled: If J is a contiguous interval of rank n and it is not the rightmost contiguous interval of that rank, then the maximum of f on J is larger than the minimum of the contiguous interval of rank $n-1$ which is closest to J from the right. Such a construction ensures that for every contiguous interval J there is k with $f^k(J) \ni 1$.

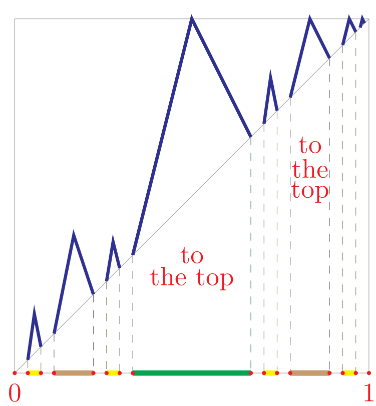


FIGURE 6. Here $s\alpha(1)$ is the middle third Cantor set.

EXAMPLE 4.8. Also it is possible to get $s\alpha(1)$ equal to the union of $\{1\}$ and a Cantor set, see Figure 7.

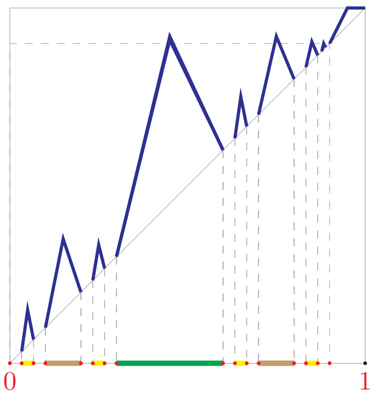


FIGURE 7. Here $s\alpha(1)$ is the union of the Cantor set and the singleton $\{1\}$.

EXAMPLE 4.9. If two special α-limit sets intersect each other, they need not be equal, see Figure 8. Compare this with properties of minimal sets.

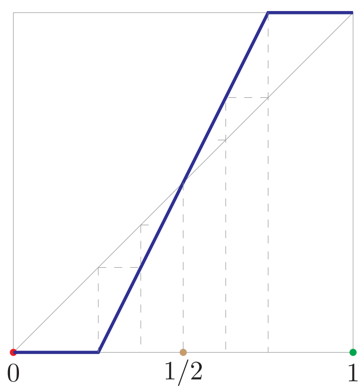


FIGURE 8. Here $s\alpha(1) = \{1/2, 1\}$ and $s\alpha(0) = \{0, 1/2\}$.

CONJECTURE 4.10. *If* $\mathrm{Int}(s\alpha(x) \cap s\alpha(y)) \neq \emptyset$ *then* $s\alpha(x) = s\alpha(y)$.

EXAMPLE 4.11. If there exists x such that $s\alpha(x) = [0, 1]$, f need not be transitive, see Figure 9.

CONJECTURE 4.12. *If there are* $x_1 \neq x_2$ *with* $s\alpha(x_1) = s\alpha(x_2) = [0, 1]$, *then* f *is transitive.*

EXAMPLE 4.13. In examples of transitive maps in Figure 10 we have $s\alpha(x) = [0, 1]$ for all x, with possible exceptions of at most three points, as claimed in Proposition 3.10.

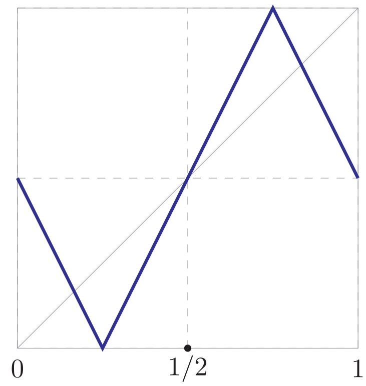


FIGURE 9. Here $s\alpha(1/2) = [0, 1]$ but f is not transitive.

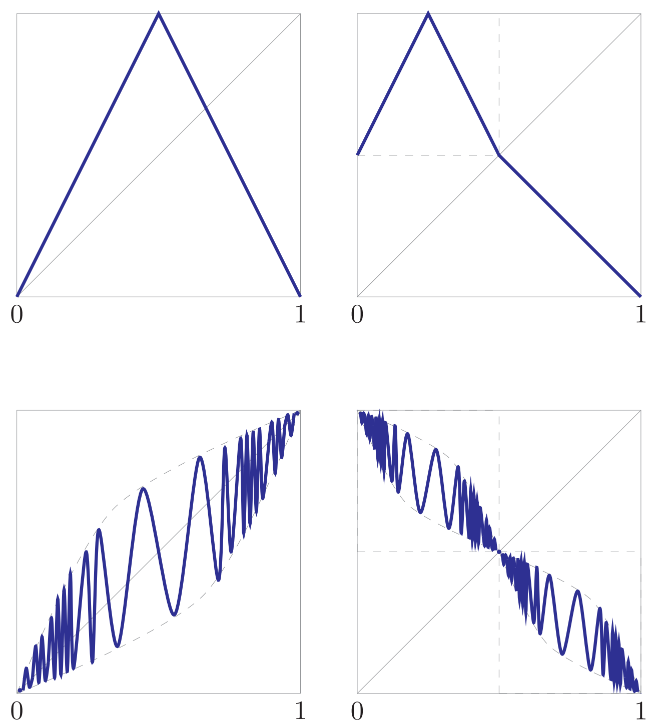


FIGURE 10. At most three points x whose $s\alpha(x)$ is not equal to $[0, 1]$.

CONJECTURE 4.14. *If* $s\alpha(x) = [0, 1]$ *then either* f *is transitive or there is* $c \in (0, 1)$ *such that* $f|_{[0,c]}$ *and* $f|_{[c,1]}$ *are transitive.*

EXAMPLE 4.15. Consider the map f from Figure 11. Then $y > 1/2$ is pre-trapped by the f-invariant interval $[0, 1/2]$. Hence by Lemma 2.8, $y \notin \bigcup_{x\in[0,1]} s\alpha(x)$.

CONJECTURE 4.16. *Any isolated point of* $s\alpha(x)$ *is periodic.*

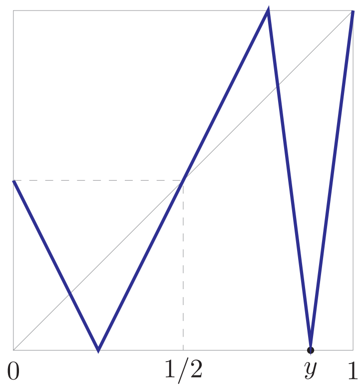


FIGURE 11. Here $y \notin \bigcup_{x\in[0,1]} s\alpha(x)$.

References

[1] S. J. Agronsky, A. M. Bruckner, J. G. Ceder, and T. L. Pearson, *The structure of ω-limit sets for continuous functions*, Real Anal. Exchange **15** (1989/90), no. 2, 483–510. MR1059418

[2] Lluís Alsedà, Jaume Llibre, and Michał Misiurewicz, *Combinatorial dynamics and entropy in dimension one*, 2nd ed., Advanced Series in Nonlinear Dynamics, vol. 5, World Scientific Publishing Co., Inc., River Edge, NJ, 2000. MR1807264

[3] Francisco Balibrea, Gabriela Dvorníková, Marek Lampart, and Piotr Oprocha, *On negative limit sets for one-dimensional dynamics*, Nonlinear Anal. **75** (2012), no. 6, 3262–3267, DOI 10.1016/j.na.2011.12.030. MR2890987

[4] Chris Bernhardt, *The ordering on permutations induced by continuous maps of the real line*, Ergodic Theory Dynam. Systems **7** (1987), no. 2, 155–160, DOI 10.1017/S0143385700003898. MR896787

[5] Ethan M. Coven and Zbigniew Nitecki, *Nonwandering sets of the powers of maps of the interval*, Ergodic Theory Dynamical Systems **1** (1981), no. 1, 9–31, DOI 10.1017/s0143385700001139. MR627784

[6] Michael W. Hero, *Special α-limit points for maps of the interval*, Proc. Amer. Math. Soc. **116** (1992), no. 4, 1015–1022, DOI 10.2307/2159483. MR1100653

[7] Sylvie Ruette, *Chaos on the interval*, University Lecture Series, vol. 67, American Mathematical Society, Providence, RI, 2017. MR3616574

[8] O. M. Šarkovsʹkiĭ, *On cycles and the structure of a continuous mapping* (Russian), Ukrain. Mat. Ž. **17** (1965), no. 3, 104–111. MR0186757

[9] O. M. Šarkovsʹkiĭ, *On a theorem of G. D. Birkhoff* (Ukrainian, with Russian and English summaries), Dopovīdī Akad. Nauk Ukraïn. RSR Ser. A **1967** (1967), 429–432. MR0212781

[10] TaiXiang Sun, HongJian Xi, and HaiLan Liang, *Special α-limit points and unilateral γ-limit points for graph maps*, Sci. China Math. **54** (2011), no. 9, 2013–2018, DOI 10.1007/s11425-011-4254-1. MR2827034

[11] Taixiang Sun, Guangwang Su, Hailan Liang, and Qiuli He, *Topological entropy and special α-limit points of graph maps*, Discrete Dyn. Nat. Soc., posted on 2011, Art. ID 132985, 7, DOI 10.1155/2011/132985. MR2794495

[12] Taixiang Sun, Yalin Tang, Guangwang Su, Hongjian Xi, and Bin Qin, *Special α-limit points and γ-limit points of a dendrite map*, Qual. Theory Dyn. Syst. **17** (2018), no. 1, 245–257, DOI 10.1007/s12346-017-0225-4. MR3776497

[13] Jin Cheng Xiong, *The attracting centre of a continuous self-map of the interval*, Ergodic Theory Dynam. Systems **8** (1988), no. 2, 205–213, DOI 10.1017/S0143385700004429. MR951269

Institute of Mathematics, NASU, Tereshchenkivs'ka 3, 01601 Kiev, Ukraine

Department of Mathematical Sciences, Indiana University - Purdue University Indianapolis, 402 N. Blackford Street, Indianapolis, IN 46202, USA
Email address: mmisiure@math.iupui.edu

Department of Mathematics, Faculty of Natural Sciences, Matej Bel University, Tajovského 40, 974 01 Banská Bystrica, Slovakia
Email address: Lubomir.Snoha@umb.sk

Contemporary Mathematics
Volume **744**, 2020
https://doi.org/10.1090/conm/744/14983

Equicontinuity of minimal sets for amenable group actions on dendrites

Enhui Shi and Xiangdong Ye

In memory of Sergiĭ Kolyada

Abstract. We show that if G is an amenable group acting on a dendrite X, then the restriction of G to any minimal set K is equicontinuous, and K is either finite or homeomorphic to the Cantor set.

1. Introduction

It is well known that every continuous action of a topological group G on a compact metric space X must have a minimal set K. A natural question is to ask what can be said about the topology of K, and the dynamics of the subsystem (K, G). The answer to this question certainly depends on the topology of X and involves the algebraic structure of G. We assume throughout that groups are topological groups, and that the actions are continuous.

In the case of an orientation-preserving group action on the circle $\mathbb{S}^1$, the topology of minimal sets and the dynamics on them are well understood. In fact, for any action of a topological group G on $\mathbb{S}^1$, the minimal set K can only be a finite set, a Cantor set, or the whole circle (see, for example, [**13**]). The interaction between the topology of K and the algebraic structure of G arises as follows.

- If K is a Cantor set, then (K, G) is semi-conjugate to a minimal action of G on $\mathbb{S}^1$.
- If $K = \mathbb{S}^1$, then (K, G) is either equicontinuous, or (K, G) is ϵ-strongly proximal for some $\epsilon > 0$, and G contains a free non-commutative subgroup (so, in particular, G cannot be amenable; see [**8**]).

The classes of minimal group actions on the circle up to topological conjugacy have been classified by Ghys using bounded Euler class (see [**4**, **5**]).

Recently, there has been considerable progress in the study of group actions on dendrites. Minimal group actions on dendrites appear naturally in the theory of 3-dimensional hyperbolic geometry (see, for example, [**2**, **10**]). Shi proved that every minimal group action on a dendrite is strongly proximal, and the acting group cannot be amenable (see [**15**, **16**]). Based on the results obtained by Marzougui and Naghmouchi in [**9**], Shi and Ye showed that an amenable group action on a dendrite

2010 *Mathematics Subject Classification.* 54H20, 37B25, 37B05, 37B40.
Key words and phrases. Equicontinuity, amenable group, minimal sets.
The work is supported by NSFC (No. 11771318, 11790274, 11431012).

always has a minimal set consisting of 1 or 2 points (see [**17**]), which is also implied by the work of Malyutin and Duchesne–Monod (see [**3**, **7**]). For group actions on dendrites with no finite orbits, Glasner and Megrelishvili showed the extreme proximality of minimal subsystems and the strong proximality of the whole system; for amenable group actions on dendrites, they showed that every infinite minimal subsystem is almost automorphic (see [**6**]). For $\mathbb{Z}$ actions on dendrites, Naghmouchi proved that every minimal set is either finite or an adding machine (see [**12**]).

We prove the following theorem in this paper, which extends the corresponding result for $\mathbb{Z}$ actions in [**12**], and answers a question proposed by Glasner and Megrelishvili in [**6**].

Theorem 1.1. *Let G be an amenable group acting on a dendrite X, and suppose that K is a minimal set for the action. Then (K, G) is equicontinuous, and K is either finite or homeomorphic to the Cantor set.*

Recently, Shi and Ye have shown that every amenable group action on uniquely arcwise connected continua (without the assumption of local connectedness) must have a minimal set consisting of 1 or 2 points (see [**18**]). We end this introduction with the following general question:

What results holding for group actions on dendrites can be extended to actions on uniquely arcwise connected continua?

In the following, we assume all the groups appearing in this paper are countable.

2. Preliminaries

2.1. Group actions. Let X be a compact metric space, Homeo(X) its homeomorphism group, and let G be a group. A group homomorphism $\phi : G \to$ Homeo(X) is called an *action* of G on X; we also write (X, G) to denote an action of G on X. For brevity, we usually write gx or $g(x)$ instead of $\phi(g)(x)$.

The *orbit* of $x \in X$ under the action of G is the set

$$Gx = \{gx \mid g \in G\}.$$

For a subset $A \subseteq X$, set $GA = \bigcup_{x\in A} Gx$; a set A is said to be *G-invariant* if $GA = A$; finally, a point $x \in X$ is called a *fixed point* of the action if $Gx = \{x\}$. If A is a G-invariant closed subset of X and $\overline{Gx} = A$ for every $x \in A$ (that is, the orbit of each point is dense), then A is called a *minimal set for the action.* In this setting every action has a minimal set by Zorn's lemma.

A Borel probability measure μ on X is called *G-invariant* if $\mu(g(A)) = \mu(A)$ for every Borel set $A \subset X$ and every $g \in G$. The following lemma follows directly from the G-invariance of the support supp(μ) (which is automatic).

Lemma 2.1. *If (X, G) is minimal and μ is a G-invariant Borel probability measure on X, then* supp$(\mu) = X$.

Lemma 2.2. *Suppose that a group G acts on a compact metric space X, and that K is a minimal set in X carrying a G-invariant Borel probability measure μ. If U and V are open sets in X such that $V \supset U$ and $g(V \cap K) \subset U \cap K$ for some $g \in G$, then $K \cap (V \setminus \overline{U}) = \emptyset$.*

Proof. Assume to the contrary that there is some $u \in K \cap (V \setminus \overline{U})$. Then there is an open neighborhood $W \ni u$ with $W \subset V \setminus \overline{U}$. By Lemma 2.1, we have $\mu(W\cap K)>0$. This then implies that $\mu(V\cap K)=\mu(g(V\cap K))\leq\mu(U\cap K)<\mu(V\cap K)$, a contradiction. □

2.2. Amenable groups. Amenability was first introduced by von Neumann. Recall that a countable group G is said to be *amenable* if there is a sequence of finite sets F_i $(i=1,2,3,\ \dots)$ such that

$$\lim_{i\to\infty}\frac{|gF_i\bigtriangleup F_i|}{|F_i|}=0$$

for every $g\in G$, where $|F_i|$ is the number of elements in F_i. The sequence (F_i) is called a *Følner sequence* and each F_i a Følner set. It is well known that solvable groups and finite groups are amenable and that any group containing a free non-commutative subgroup is not amenable. One may consult the monograph of Paterson [**14**] for the proofs of the following lemmas.

Lemma 2.3. *Every subgroup of an amenable group is amenable.*

Lemma 2.4. *A group G is amenable if and only if every action of G on a compact metric space X has a G-invariant Borel probability measure on X.*

2.3. Dendrites. A *continuum* is a non-empty connected compact metric space. A continuum is said to be *non-degenerate* if it is not a single point. An *arc* is a continuum which is homeomorphic to the closed interval $[0,1]$. A continuum X is *uniquely arcwise connected* if for any two points $x\neq y\in X$ there is a unique arc $[x,y]$ in X connecting x and y.

A *dendrite* X is a locally connected, uniquely arcwise connected, continuum. If Y is a subcontinuum of a dendrite X, then Y is called a *subdendrite* of X. For a dendrite X and a point $c\in X$, if $X\setminus\{c\}$ is not connected, then c is called a *cut point* of X; if $X\setminus\{c\}$ has at least 3 components, then c is called a *branch point* of X.

Lemmas 2.5 to 2.8 are taken from [**11**].

Lemma 2.5. *Let X be a dendrite with metric d. Then, for every $\epsilon>0$, there is a $\delta>0$ such that* $\operatorname{diam}([x,y])<\epsilon$ *whenever $d(x,y)<\delta$.*

Lemma 2.6. *Let X be a dendrite. If A_i $(i=1,2,3,\dots)$ is a sequence of mutually disjoint sub-dendrites of X, then* $\operatorname{diam}(A_i)\to 0$ *as $i\to\infty$.*

Lemma 2.7. *Let X be a dendrite. Then X has at most countably many branch points. If X is nondegenerate, then the cut point set of X is uncountable.*

Lemma 2.8. *Let X be a dendrite and $c\in X$. Then each component U of $X\setminus\{c\}$ is open in X, and $\overline{U}=U\cup\{c\}$.*

Now we give a proof of the following technical lemma.

Lemma 2.9. *Let X be a dendrite and let $f:X\to X$ be a homeomorphism. Suppose o is a fixed point of f, and let c_1,c_2 be cut points of X different from o. Suppose that U is a component of $X\setminus\{c_1\}$ not containing o, that V is a component of $X\setminus\{c_2\}$ not containing o, and that $f(c_1)\in V$. Then $f(U)\subset V$.*

Proof. Assume to the contrary that there is some $u\in U$ with $f(u)\notin V$. Since c_2 is a cut point, $f(c_1)\in V$, $o\notin V$, and $f(o)=o$, we have $c_2\in[f(o),f(c_1)]$ and $c_2\in[f(u),f(c_1)]$. This implies that $f^{-1}(c_2)\in[o,c_1]\cap[u,c_1]=\{c_1\}$ since $o\notin U$. Thus $f(c_1)=c_2$, which contradicts the assumption that $f(c_1)\in V$. □

If $[a, b]$ is an arc in a dendrite X, denote by $[a, b)$, $(a, b]$, and (a, b) the sets $[a, b] \setminus \{b\}$, $[a, b] \setminus \{a\}$, and $[a, b] \setminus \{a, b\}$, respectively.

2.4. Equicontinuity. Let X be a compact metric space with metric d, and let G be a group acting on X. Two points $x, y \in X$ are said to be *regionally proximal* if there are sequences (x_i), (y_i) in X and (g_i) in G such that $x_i \to x$ and $y_i \to y$ as $i \to \infty$, and $\lim g_i x_i = \lim g_i y_i = w$ for some $w \in X$. If x, y are regionally proximal and $x \neq y$, then $\{x, y\}$ is said to be a *non-trivial regionally proximal pair.* The action (X, G) is *equicontinuous* if, for every $\epsilon > 0$, there is a $\delta > 0$ such that $d(gx, gy) < \epsilon$ for all $g \in G$ whenever $d(x, y) < \delta$.

The following lemma can be found in [**1**].

LEMMA 2.10. *Suppose* (X, G) *is a group action. Then* (X, G) *is equicontinuous if and only if it contains no non-trivial regionally proximal pair.*

3. Proof of the main theorem

In this section we are going to show our main result. Before doing this we state two simple lemmas.

LEMMA 3.1. *Suppose a group* G *acts on the closed interval* $[0, 1]$. *If* $K \subset [0, 1]$ *is minimal, then* K *contains at most* 2 *points.*

PROOF. Let $x = \inf K$ and $y = \sup K$. Then G preserves the set $\{x, y\}$, so $K = \{x, y\}$ by the minimality of K. $\square$

LEMMA 3.2 (See [**17**]). *Let* G *be an amenable group acting on a dendrite* X. *Then there is a* G*-invariant set consisting of* 1 *or* 2 *points.*

Now we are ready to prove the main result.

PROOF OF THEOREM 1.1. We first show that (K, G) is equicontinuous.

Assume to the contrary that (K, G) is not equicontinuous. Then by Lemma 2.10, there are $u \neq v \in K$ such that u, v are regionally proximal; that is, there are sequences $(u_i), (v_i)$ in X and (g_i) in G with

$$u_i \to u, v_i \to v, \ \lim g_i x_i = \lim g_i y_i = w \tag{3.1}$$

as $i \to \infty$ for some $w \in K$.

By Lemma 3.2, there are $o_1, o_2 \in X$ such that $\{o_1, o_2\}$ is a G-invariant set. Then $[o_1, o_2]$ is G-invariant by the unique arcwise connectedness of X. From the assumption, K is infinite so $K \cap [o_1, o_2] = \emptyset$ by Lemma 3.1. Without loss of generality, we may suppose that $o_1 = o_2$ and denote this common point by o; otherwise, we need only collapse $[o_1, o_2]$ to one point. Then o is a fixed point for the action.

Case 1. $[u, o] \cap [v, o] = \{o\}$ (see Fig.1(1)). By Lemma 2.7, we can choose cut points $c_1 \in (u, o)$ and $c_2 \in (v, o)$. Let D_u be the component of $X \setminus \{c_1\}$, which contains u; let D_v be the component of $X \setminus \{c_2\}$, which contains v. From minimality and Lemma 2.8, there is some $g' \in G$ with $g'w \in D_u$. From (3.1) and Lemma 2.5, we have

$$u_i \in D_u, v_i \in D_v \text{ and } g' g_i [u_i, v_i] \subset D_u \tag{3.2}$$

for large enough i. Write $g = g' g_i$. Then $o \in [u_i, v_i]$ and $g(o) \in D_u$. This is a contradiction, since o is fixed by G.

Case 2. $[u,o]\cap[v,o]=[z,o]$ for some $z\neq o$.

Subcase 2.1. $z=v$ (see Fig.1(2)). Then $u\neq z$ and $z\in K$. Take a cut point $c_1\in(u,z)$ and let D_u be the component of $X\setminus\{c_1\}$ which contains u. Then $v\notin D_u$, and there is some $g\in G$ with $gz\in D_u$ by the minimality of K. Take a cut point $c_2\in(z,o)$ which is sufficiently close to z to ensure that $g(c_2)\in D_u$. Let D_z be the component of $X\setminus\{c_2\}$ which contains z. By Lemma 2.4, there is a G-invariant Borel probability measure on K. Applying Lemma 2.9, we get $g(D_z)\subset D_u$, which contradicts Lemma 2.2, since $z\in D_z\setminus\overline{D}_u$.

Subcase 2.2. $z=u$. In this case we can deduce a contradiction along the lines of the argument in Subcase 2.1.

Subcase 2.3. $z\neq u$ and $z\neq v$ (see Fig.1(3)). Take a cut point $c_1\in(u,z)$. Let D_u be the component of $X\setminus\{c_1\}$, which contains u. Similar to the argument in Case 1, there is some $g\in G$ with $g(z)\in D_u$. Take a cut point $c_2\in(z,o)$ which is sufficiently close to z to ensure that $g(c_2)\in D_u$. Let D_z be the component of $X\setminus\{c_2\}$, which contains z. Then $g(D_z)\subset D_u$ by Lemma 2.9. This contradicts Lemma 2.2 since $v\in D_z\setminus\overline{D}_u$.

Now we prove that if K is not finite, then K is homeomorphic to the Cantor set. If not, then there is some non-degenerate connected component Y of K. Clearly, for any $g,g'\in G$, either $g(Y)=g'(Y)$ or $g(Y)\cap g'(Y)=\emptyset$. This, together with Lemma 2.6 and the equicontinuity of (K,G), implies that the subgroup $H=\{g\in G: g(Y)=Y\}$ has finite index in G. It follows that (Y,H) is minimal. This contradicts Lemma 3.2 and Lemma 2.3, since Y is a non-degenerate dendrite. $\square$

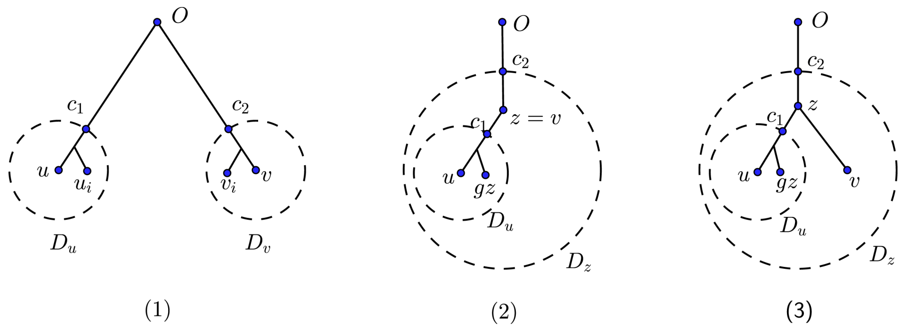


Fig. 1

Acknowledgment. The authors would like to thank Eli Glasner for sending us the early version of his work with Megrelishvili.

References

[1] Joseph Auslander, *Minimal flows and their extensions*, North-Holland Mathematics Studies, vol. 153, North-Holland Publishing Co., Amsterdam, 1988. Notas de Matemática [Mathematical Notes], 122. MR956049

[2] Brian H. Bowditch, *Hausdorff dimension and dendritic limit sets*, Math. Ann. **332** (2005), no. 3, 667–676, DOI 10.1007/s00208-005-0647-1. MR2181766

[3] Bruno Duchesne and Nicolas Monod, *Group actions on dendrites and curves* (English, with English and French summaries), Ann. Inst. Fourier (Grenoble) **68** (2018), no. 5, 2277–2309. MR3893770

[4] Étienne Ghys, *Groupes d'homéomorphismes du cercle et cohomologie bornée* (French, with English summary), The Lefschetz centennial conference, Part III (Mexico City, 1984), Contemp. Math., vol. 58, Amer. Math. Soc., Providence, RI, 1987, pp. 81–106. MR893858

[5] Étienne Ghys, *Groups acting on the circle*, Enseign. Math. (2) **47** (2001), no. 3-4, 329–407. MR1876932

[6] E. Glasner and M. Megrelishvili. Group actions on treelike compact spaces. *Sci. China Math.* to appear.

[7] A. V. Malyutin, *On groups acting on dendrons* (Russian, with English summary), Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) **415** (2013), no. Geometriya i Topologiya. 12, 62–74, DOI 10.1007/s10958-016-2688-2; English transl., J. Math. Sci. (N.Y.) **212** (2016), no. 5, 558–565. MR3473016

[8] Gregory Margulis, *Free subgroups of the homeomorphism group of the circle* (English, with English and French summaries), C. R. Acad. Sci. Paris Sér. I Math. **331** (2000), no. 9, 669–674, DOI 10.1016/S0764-4442(00)01694-3. MR1797749

[9] Habib Marzougui and Issam Naghmouchi, *Minimal sets for group actions on dendrites*, Proc. Amer. Math. Soc. **144** (2016), no. 10, 4413–4425, DOI 10.1090/proc/13103. MR3531191

[10] Yair N. Minsky, *On rigidity, limit sets, and end invariants of hyperbolic 3-manifolds*, J. Amer. Math. Soc. **7** (1994), no. 3, 539–588, DOI 10.2307/2152785. MR1257060

[11] Sam B. Nadler Jr., *Continuum theory*, Monographs and Textbooks in Pure and Applied Mathematics, vol. 158, Marcel Dekker, Inc., New York, 1992. An introduction. MR1192552

[12] Issam Naghmouchi, *Dynamical properties of monotone dendrite maps*, Topology Appl. **159** (2012), no. 1, 144–149, DOI 10.1016/j.topol.2011.08.020. MR2852955

[13] Andrés Navas, *Grupos de difeomorfismos del círculo* (Spanish, with English summary), Ensaios Matemáticos [Mathematical Surveys], vol. 13, Sociedade Brasileira de Matemática, Rio de Janeiro, 2007. MR2394157

[14] Alan L. T. Paterson, *Amenability*, Mathematical Surveys and Monographs, vol. 29, American Mathematical Society, Providence, RI, 1988. MR961261

[15] Enhui Shi, *Free subgroups of dendrite homeomorphism group*, Topology Appl. **159** (2012), no. 10-11, 2662–2668, DOI 10.1016/j.topol.2012.04.001. MR2923436

[16] Enhui Shi, Suhua Wang, and Lizhen Zhou, *Minimal group actions on dendrites*, Proc. Amer. Math. Soc. **138** (2010), no. 1, 217–223, DOI 10.1090/S0002-9939-09-10000-X. MR2550186

[17] Enhui Shi and Xiangdong Ye, *Periodic points for amenable group actions on dendrites*, Proc. Amer. Math. Soc. **145** (2017), no. 1, 177–184, DOI 10.1090/proc/13206. MR3565370

[18] E. H. Shi and X. Ye. Periodic points for amenable group actions on uniquely arcwise connected continua. arXiv:1710.03411.

School of Mathematical Sciences, Soochow University, Suzhou 215006, P. R. China
Email address: ehshi@suda.edu.cn

Wu Wen-Tsun Key Laboratory of Mathematics, USTC, Chinese Academy of Sciences and Department of Mathematics, University of Science and Technology of China, Hefei, Anhui 230026, China
Email address: yexd@ustc.edu.cn

Contemporary Mathematics
Volume **744**, 2020
https://doi.org/10.1090/conm/744/14985

On weak rigidity and weakly mixing enveloping semigroups

Ethan Akin, Eli Glasner, and Benjamin Weiss

In memory of Sergii Kolyada, colleague and friend

Abstract. The question we deal with here, which was presented to us by Joe Auslander and Anima Nagar, is whether there is a nontrivial cascade (X,T) whose enveloping semigroup, as a dynamical system, is topologically weakly mixing (WM). After an introductory section recalling some definitions and classic results, we establish some necessary conditions for this to happen, and in the final section we show, using Ratner's theory, that the enveloping semigroup of the 'time one map' of a classical horocycle flow is weakly mixing.

1. Introduction

A *cascade* is a homeomorphism T on a compact Hausdorff space X. We call the system (X,T) metric when X is metrizable. If A is a closed, invariant subset of X, i.e. $T(A) = A$, then the restriction of T to A defines the *subsystem* on A. If (X_1,T_1) is a cascade and $\pi : X \to X_1$ is a continuous surjection such that $\pi \circ T = T_1 \circ \pi$ then π is a surjective *cascade morphism* and (X_1,T_1) is a *factor* of (X,T).

For $A, B \subset X$ we let $N(A,B) = \{i \in \mathbb{Z} : A \cap T^{-i}(B) \neq \emptyset\}$. Here $\mathbb{Z}$ denotes the set of integers and we use $\mathbb{N}$ for the set of non-negative integers. We write $N(x,B)$ for $N(\{x\},B)$ when $x \in X$. The cascade is *transitive* when $N(U,V)$ is nonempty for every pair of nonempty, open subsets U,V of X. We let $\mathcal{O}(x) = \{T^n x : n \in \mathbb{Z}\}$ denote the *orbit* of x. A point x is a *transitive point* when the orbit-closure $\overline{\mathcal{O}(x)} = X$. When (X,T) admits transitive points the system is called *point transitive*. A point transitive system is transitive and the converse holds when the system is metric. In a metric transitive system the set of transitive points forms a dense G_δ set by the Baire Category Theorem . The system (X,T) is called *totally transitive* if for every $0 \neq n \in \mathbb{Z}$ the system (X,T^n) is transitive.

When the homeomorphism is understood, we will refer to the system X.

The system X is called *minimal* when every point of X is a transitive point, or, equivalently, when X contains no proper, closed, invariant subset. A point $x \in X$ is called a *minimal point*, or an *almost periodic point* (hereafter abbreviated a.p.) when the restriction of T to $\overline{\mathcal{O}(x)}$ is minimal.

A subset of $\mathbb{Z}$ is called *thick* when it contains blocks of consecutive integers of arbitrary length. It is called *syndetic* when it meets every thick subset. A point x

2010 *Mathematics Subject Classification.* Primary 37B05, 37A17, 54H20.
Key words and phrases. Enveloping semigroup, Weak rigidity, Weak mixing, Horocyclic flow.

is minimal if and only if $N(x,U)$ is syndetic for every neighborhood U of x, see, e.g. [**Aus**] Theorem 1.8.

An *ambit* (X, x_0, T) is a point transitive cascade (X,T) with a chosen transitive base point x_0. If $\pi : (X,T) \to (X_1, T_1)$ is a surjective cascade morphism and (X, x_0, T) is an ambit then $(X_1, \pi(x_0), T_1)$ is an ambit factor.

The enveloping semigroup $E(X,T)$ (or simply $E(X)$) is the closure in X^X of the set $\{T^i : i \in \mathbb{Z}\}$. If T_* denotes composition with T then $(E(X), \mathrm{Id_X}, \mathrm{T_*})$ is an ambit, where $\mathrm{Id_X} = \mathrm{T}^0$ is the identity map. If $x \in X$ then the evaluation map $p \mapsto px$, is a surjection of ambits $ev_x : (E(X), \mathrm{Id}_X, T_*) \to (\overline{\mathcal{O}(x)}, x, T)$ taking the enveloping semigroup onto the orbit closure of x.

Proposition 1.1. *For any non-empty index set I and the product system $(X^I, T^{(I)})$ we can identify $E(X)$ with $E(X^I)$ by mapping T^i to $(T^{(I)})^i$. Hence, for any k-tuple $(x_1, x_2, \dots, x_k) \in X^k$ the ambit $\overline{\mathcal{O}(x_1, x_2, \dots, x_k)}$ is a factor of $E(X)$ via evaluation at $(x_1, x_2, \dots, x_k)$, i.e. by $p \mapsto (px_1, px_2, \dots, px_k)$ for $p \in E(X)$. In addition, $E(X)$ can be expressed as the inverse limit of these factors.*

Proof. The first two claims are easy to check. To prove the last one, observe that $(E(X), T_*)$ is the orbit closure $(\overline{\mathcal{O}(Id_X)}, T^X)$ in (X^X, T^X). Partially ordered by inclusion, the collection of finite (unordered) subsets of X forms a directed set with respect to which $\overline{\mathcal{O}(Id_X)}$ is the inverse limit $\lim_{\leftarrow} \overline{\mathcal{O}(x_1, x_2, \dots, x_k)}$ via the evaluation maps. □

If A is closed invariant subset of X then the restriction map $E(X) \to E(A)$ is a surjective morphism. A surjective morphism $\pi : (X,T) \to (X_1, T_1)$ induces a surjective morphism $\pi_* : E(X,T) \to E(X_1, T_1)$ by mapping T^i to T_1^i.

2. Recurrence

For the system (X,T) and the point $x \in X$ the limit point sets $\alpha(x)$ and $\omega(x)$ are the sets of limit points of the bi-infinite sequence $\{T^i(x) : i \in \mathbb{Z}\}$ as i tends to $-\infty$ and $+\infty$ respectively. Both $\alpha(x)$ and $\omega(x)$ are non-empty, closed, invariant subsets. The orbit-closure of x consists of the orbit and $\alpha(x) \cup \omega(x)$.

For $(E(X), T_*)$ we denote

$$Ad_+(X) = \bigcap \mathrm{cls}\, \{T^n : n \geq 1\}, \quad Ad_-(X) = \bigcap \mathrm{cls}\, \{T^n : n \leq -1\}$$

and $Ad(X) = Ad_+(X) \cup Ad_-(X)$. $Ad(X)$ is called the *adherence semigroup* of (X,T). It is a closed subsemigroup of $E(X)$. Furthermore, $ev_x(Ad_-(X)) = \alpha(x)$ and $ev_x(Ad_+(X)) = \omega(x)$.

A point x is called *recurrent* if $x \in \alpha(x) \cup \omega(x)$ and *positive recurrent* if $x \in \omega(x)$. Thus, x is recurrent (or positive recurrent) if $x = px$ for some $p \in Ad(X)$ (resp. for some $p \in Ad_+(X)$).

A point is recurrent if and only if $N(x,U)$ is infinite for every neighborhood U of x and it is positive recurrent if for every such U, $N(x,U) \cap \mathbb{N}$ is infinite. If $\pi : (X,T) \to (X_1, T_1)$ is a surjective morphism and x is a recurrent point in X then $\pi(x)$ is a recurrent point in X_1.

Lemma 2.1. *If (X,T) is the inverse limit of a system $\{(X_i, T_i)\}$ indexed by the directed set I and with $\pi_i : X \to X_i$ the corresponding projections, then $x \in X$ is recurrent (or positive recurrent) if and only if $\pi_i(x)$ is recurrent (resp. positive recurrent) for every $i \in I$.*

Proof. If $px = x$ for some $p \in Ad(X)$ then $p\pi_i(x) = \pi_i(px) = \pi_i(x)$ for all i.

Conversely, if $\pi_i(x)$ is recurrent, then $K_i = \{p \in Ad(X) : p\pi_i(x) = \pi_i(x)\}$ is a closed, non-empty compact subset of $Ad(X)$. Since I is directed, these sets form a filterbase and so by compactness the intersection K is a nonempty subset of $Ad(X)$ which consists of the $p \in Ad(X)$ such that $px = x$.

For positive recurrence apply the same argument using Ad_+ instead. □

The system (X, T) is called *pointwise recurrent* when every point of X is recurrent, i.e. $x \in \alpha(x) \cup \omega(x)$ for all $x \in X$. It is called *pointwise forward recurrent* when $x \in \omega(x)$ for all $x \in X$.

It is clear that subsystems and factors of a system which is pointwise recurrent or pointwise forward recurrent satisfy the corresponding property. By Lemma 2.1 these properties are preserved by inverse limits as well.

Lemma 2.2. *If the product system $X \times X$ is pointwise recurrent then either (X, T) or (X, T^{-1}) (or both) is pointwise forward recurrent.*

Proof. If neither (X, T) nor (X, T^{-1}) is pointwise forward recurrent, then there exist $x, y \in X$ such that $x \notin \omega(x)$ and $y \notin \alpha(y)$. Since $\omega(x, y) \subset \omega(x) \times \omega(y)$ and similarly for α it follows that $(x, y) \notin \alpha(x, y) \cup \omega(x, y)$. □

Definition 2.3. Given a cascade (X, T), let A be an invariant subset of X.

(a) We say that A is an *isolated invariant set* if there is a closed subset $U \subset X$ containing A in its interior such that A is the maximum closed invariant subset of U, i.e. if a closed invariant set B is contained in U then $B \subset A$; or, equivalently if $A = \bigcap_{n\in\mathbb{Z}} T^n(U)$. The set U is called an *isolating neighborhood* for A.

(b) We will call A an *attractor* if there is a closed subset $U \subset X$ containing A in its interior such that $A = \bigcap_{n\in\mathbb{N}} T^n(U)$.

Thus, an attractor is isolated and an isolated invariant set is closed.

The above is not the standard definition of an attractor but is equivalent to it, see [**Ak-93**] Theorem 3.3(b). If A is an attractor for (X, T) then there exists a *dual repeller* R, i.e. an attractor for (X, T^{-1}), disjoint from A and such that for all $x \in X \setminus (A \cup R)$, $\omega(x) \subset A$ and $\alpha(x) \subset R$. In particular, no point of $X \setminus (A \cup R)$ is recurrent. See [**Ak-93**] Proposition 3.9. From this we prove the following result from [**Ak-96**].

Theorem 2.4. *If (X, T) is pointwise forward recurrent and A is an isolated invariant subset of X then A is clopen in X.*

Proof. Let U be an isolating neighborhood for A. If $x \in \bigcap_{n\in\mathbb{N}} T^{-n}(U)$ then $T^n(x) \in U$ for all $n \in \mathbb{N}$. Since U is closed, $\omega(x) \subset U$. Since $\omega(x)$ is invariant, it is a subset of A. Because the system is pointwise forward recurrent, $x \in \omega(x)$ and so $x \in A$. Thus, $A = \bigcap_{n\in\mathbb{N}} T^{-n}(U)$. Thus, A is an attractor for (X, T^{-1}) with dual repeller R. Since no point of $X \setminus (A \cup R)$ is recurrent, this set must be empty. Thus, $A = X \setminus R$ is open. □

Examples 2.5. (a) Let $L = IP\{10^t\}_{t=1}^{\infty} \subset \mathbb{N}$ be the IP-sequence generated by the powers of ten, i.e.

$$L = \{10^{a_1} + 10^{a_2} + \cdots + 10^{a_k} : 1 \leq a_1 < a_2 < \cdots < a_k\}.$$

Let $f = 1_L$ and let $X = \overline{\mathcal{O}_T(f)} \subset \{0,1\}^{\mathbb{Z}}$, where T is the shift on $\Omega = \{0,1\}^{\mathbb{Z}}$. It is easy to check that X is a Cantor set with a single fixed point $\mathbf{0}$ and such that every other orbit is dense, so that (X,T) is pointwise recurrent. The fixed point $\mathbf{0}$ forms an isolated invariant set which is not clopen and it follows that neither (X,T) nor (X,T^{-1}) is pointwise forward recurrent, so $(X \times X, T \times T)$ is not pointwise recurrent. In fact, the point f is clearly forward but not backward recurrent and there are points in X which are backward but not forward recurrent. (See also [**DY**] and [**Ak-16**, Theorem 4.16].)

(b) As was shown in [**Ak-93**, page 180] the stopped torus example (X,T) is a connected, topologically mixing, pointwise recurrent system, containing a unique fixed point as its mincenter [1]. Furthermore, the fixed point is a proper isolated closed invariant subset. So again neither (X,T) nor (X,T^{-1}) is pointwise forward recurrent.

(c) Let (X,T) be an infinite subshift. Such a system is *expansive* which says exactly that the diagonal Δ_X is an isolated, invariant subset of $X \times X$. Since X is infinite, the diagonal is not clopen and so neither $(X \times X, T \times T)$ nor $(X \times X, (T \times T)^{-1})$ is pointwise forward recurrent. Hence, the product system on X^4 is not pointwise recurrent. On the other hand, if, for example, (X,T) is minimal then it and its inverse are pointwise forward recurrent.

(d) A minimal system (X,T) is called *doubly minimal* if every point in

$$X \times X \setminus \bigcup \{(\mathrm{Id} \times T)^n \Delta_X : n \in \mathbb{Z}\}$$

has a dense orbit in $X \times X$. Such systems exist in abundance, and many of them are also subshifts, hence expansive (see [**W**]). It follows directly from the definition that if X is a doubly minimal subshift, then $X \times X$ is pointwise recurrent. Thus such a system (X,T) has the property that $X \times X$ is pointwise recurrent but as in (a) neither the subshift $(X \times X, T \times T)$ nor its inverse can be pointwise forward recurrent. Hence, X^4 is not pointwise recurrent.

The system (X,T) is called *weakly rigid* (see [**GMa**]) when $\mathrm{Id_X}$ is a recurrent point of $E(X)$. That is, $Id_X \in Ad(X)$ and so either $Id_X \in Ad_+(X)$ or $Id_X \in Ad_-(X)$.

THEOREM 2.6. *For a cascade (X,T) the following are equivalent.*

(i) *The system (X,T) is weakly rigid.*
(ii) $E(X) = Ad(X)$.
(iii) *The product system $(X^I, T^{(I)})$ is weakly rigid for every nonempty index set I.*
(iv) *The product system $(X^I, T^{(I)})$ is pointwise recurrent for every nonempty index set I.*
(v) $(E(X), T_*)$ *is pointwise recurrent.*
(vi) *For any k-tuple $(x_1, x_2, \dots, x_k) \in X^k$ the subsystem $\overline{\mathcal{O}(x_1, x_2, \dots, x_k)}$ is pointwise recurrent.*
(vii) *The collection $\{N(x,U) : x \in X, U$ an open set with $x \in U\}$ forms a filter base.*

PROOF. (i) $\Leftrightarrow$ (ii): As it is the orbit closure of Id_X, $E(X)$ is contained in the closed, invariant set $Ad(X)$ if and only if $Id_X \in Ad(X)$.

[1] The mincenter of a dynamical system is the closure of the union of its minimal subsystems.

(i) ⇔ (iii): $E(X,T) \cong E(X^I, T^I)$.

(i) ⇒ (vi): The evaluation map at $(x_1, \dots, x_k)$ takes Id_X to $(x_1, \dots, x_k)$. Similarly, (iii) ⇒ (iv).

(iv) ⇒ (v): $E(X)$ is a subsystem of X^X.

(v) ⇒ (i): Obvious.

(vi) ⇒ (v): $E(X)$ is the inverse limit of the $\overline{\mathcal{O}(x_1, x_2, \dots, x_k)}$'s.

(vi) ⇔ (vii): If U_i is an open set containing x_i for $i = 1, \dots, k$, then $N((x_1, \dots, x_k), U_1 \times \dots U_k) = \bigcap_{i=1}^k N(x_i, U_i)$. □

It follows that weak rigidity is preserved by factors, subsystems and inverse limits.

When $Id_X \in Ad_+(X)$ then $E(X) = Ad_+(X)$ and every $(X^I, T^{(I)})$ is pointwise forward recurrent. Conversely, if every $\overline{\mathcal{O}(x_1, x_2, \dots, x_k)}$ is pointwise forward recurrent then $Id_X \in Ad_+(X)$. In particular, if X is weakly rigid then either (X,T) or (X, T^{-1}) is pointwise forward recurrent.

Recall that $x \in X$ is almost periodic, or a.p., when it has a minimal orbit closure. We call the system (X,T) *pointwise almost periodic* or pointwise a.p. when every point $x \in X$ is a.p. A point transitive, pointwise a.p. system is minimal.

We will call the system *distal* when $(X \times X, T \times T)$ is pointwise a.p. Again this is not the usual definition but the equivalence is described in [**E**, Theorem 1] (see also [**Aus**] Theorem 5.6). Distality is preserved by factors, subsystems and inverse limits. Using [**E**] Theorem 1 or [**Aus**] Theorem 5.6 and arguments similar to those of Theorem 2.6 one can prove the following.

THEOREM 2.7. *For a cascade (X,T) the following are equivalent.*

(i) *The system (X,T) is distal.*
(ii) *$(E(X), T_*)$ is a minimal system.*
(iii) *$E(X)$ is a group.*
(iv) *The product system $(X^I, T^{(I)})$ is pointwise a.p. for every nonempty index set I.*
(v) *For any k-tuple $(x_1, x_2, \dots, x_k) \in X^k$ the subsystem $\overline{\mathcal{O}(x_1, x_2, \dots, x_k)}$ is pointwise a.p.*
(vi) *There exists a filter $\mathcal{F}$ of syndetic sets such that for every $x \in X$ and open subset U containing x, the return time set $N(x,U)$ is an element of $\mathcal{F}$.*

COROLLARY 2.8. *If system (X,T) is distal, then it is weakly rigid. Moreover, $E(X) = Ad_+(X) = Ad_-(X)$.*

PROOF. Since $E(X)$ is minimal, it is equal to each of the nonempty, closed invariant subsets $Ad_+(X)$ and $Ad_-(X)$. In particular, $Id_X \in E(X)$ lies in $Ad_+(X)$ and $Ad_-(X)$. □

The system X is *weak mixing* (hereafter WM) when the product system $X \times X$ is transitive. The transitivity property is preserved by factors and inverse limits and so the same is true for WM.

Recall that subset of $\mathbb{Z}$ is called *thick* when it contains runs of arbitrary length. The following is a classic result of Furstenberg (see [**F**, Proposition II.3]).

THEOREM 2.9. *For a cascade (X,T) the following are equivalent.*

(i) *The system (X,T) is WM.*
(ii) *The product system $(X^I, T^{(I)})$ is WM for every nonempty index set I.*

(iii) *For every pair U, V of nonempty open subsets of X, the visiting time set $N(U, V)$ is thick.*
(iv) *The system (X, T) is transitive and for every nonempty open subset U of X, the return time set $N(U, U)$ is thick.*
(v) *The collection $\{N(U, V) : U, V$ non-empty open sets$\}$ forms a filter base.*

Corollary 2.10. *If a cascade (X, T) is WM, then it it totally transitive.*

Proof. Every $N(U, V)$ is thick and so meets $k\mathbb{Z}$ for any positive integer k. □

Corollary 2.11. *Let (X, T) be a nontrivial, WM system. Then every transitive point in X is recurrent.*

Proof. Let $x \in X$ be a transitive point. If x is not recurrent, then it has a neighborhood U such that $N(x, U)$ is finite. As x is not recurrent, it is not periodic and so we can remove a finite set to obtain a neighborhood U' such that $N(x, U') = \{0\}$. As $N(x, U' \setminus \{x\}) = \emptyset$ and x is a transitive point it follows that $U' \setminus \{x\} = \emptyset$. That is, the singleton set $\{x\}$ is clopen. Now, in this case, the sets $U = \{(x, x)\}$ and $V = \{(x, Tx)\}$ are clopen subsets of $X \times X$ but there is no $k \in \mathbb{Z}$ such that $(T \times T)^k U \cap V \neq \emptyset$. Thus, (X, T) is not weak mixing. □

3. Some obstructions to WM of $E(X, T)$

Recall that an ambit (X, x_0, T) is an enveloping semigroup if and only if it is *point universal*; i.e. it satisfies the following condition: For every $x \in X$ there is a (unique) homomorphism of ambits $(X, x_0, T) \to (X, x, T)$ (see e.g. [**GMe**, Proposition 2.6]).

We will say that X *has a WM enveloping semigroup* when the system $(E(X), T_*)$ is WM.

Call a subset F of $\mathbb{Z}$ *diff-thick* when the difference set $\{i - j : i, j \in F\}$ is thick.

Theorem 3.1. *For a cascade (X, T) the following are equivalent.*

(i) *The system (X, T) has a WM enveloping semigroup.*
(ii) *For any k-tuple $(x_1, x_2, \ldots, x_k) \in X^k$ the ambit $\overline{\mathcal{O}(x_1, x_2, \ldots, x_k)}$ is WM.*
(iii) *There exists a filter $\mathcal{F}$ of diff-thick sets such that for every $x \in X$ and open subset U containing x, the return time set $N(x, U)$ is an element of $\mathcal{F}$.*

Proof. (i) $\Leftrightarrow$ (ii): Each ambit $\overline{\mathcal{O}(x_1, x_2, \ldots, x_k)}$ is a factor of $E(X)$ and $E(X)$ is an inverse limit of these factors. WM is preserved by factors and inverse limits.

(iii) $\Rightarrow$ (ii): If U_1, U_2 are open sets which meet $\overline{\mathcal{O}(z)}$ with $z = (x_1, x_2, \ldots, x_k)$ then there exist $i_1, i_2 \in \mathbb{Z}$ such that $(T^{(k)})^{i_1}(z) \in U_1, (T^{(k)})^{i_2}(z) \in U_2$. Let $U = (T^{(k)})^{-i_1}(U_1) \cap (T^{(k)})^{-i_2}(U_2)$. Since $\mathcal{F}$ is a filter, $N(z, U) \in \mathcal{F}$. If $(T^{(k)})^{k_1}(z)$, $(T^{(k)})^{k_2}(z) \in U$ then $k_2 - k_1 + (i_2 - i_1) \in N(U_1, U_2)$. Since $N(z, U)$ is diff-thick and the translate of a thick set is thick, $N(U_1, U_2)$ is thick. Hence, $\overline{\mathcal{O}(z)}$ is WM.

(ii) $\Rightarrow$ (iii): If U_ℓ is a neighborhood of x_ℓ for $\ell = 1, \ldots, k$ and $U = U_1 \times \cdots \times U_k$ then U is a neighborhood of $z = (x_1, x_2, \ldots, x_k)$ and so $N(U, U)$ is thick. Furthermore, as above, $N(U, U) = N(z, U) - N(z, U)$ and so $N(z, U)$ is diff-thick. Since $N(z, U) = \bigcap_{\ell=1}^k N(x_\ell, U_\ell)$, it follows that $\{N(x, V) : x \in X, V$ open in X with $x \in V\}$ generates a filter of diff-thick sets.

□

Corollary 3.2. *If system (X,T) has a WM enveloping semigroup then it is weakly rigid.*

Proof. Corollary 2.11 and condition (ii) of Theorem 3.1 imply condition (vi) of Theorem 2.6.

□

Theorem 3.3. *Let (X,T) be a transitive cascade.*

(1) *If (X,T) is weakly rigid, or, more generally, if $(X \times X, T \times T)$ is pointwise recurrent, then X does not admit a proper isolated, closed, invariant subset.*
(2) *If (X,T) is weakly rigid and WM, then X is connected.*

Proof. If (X,T) is weakly rigid, then $(X\times X, T\times T)$ is pointwise recurrent. By Lemma 2.2 either (X,T) or (X,T^{-1}) is pointwise forward recurrent. By Theorem 2.4 an isolated, closed, invariant proper subset would be clopen, contradicting the transitivity of X.

Now suppose that X is not connected and so there exist U, V proper, disjoint clopen sets with union X. Then $W = (U \times U) \cup (V \times V)$ is a proper, clopen subset of $X \times X$ containing the diagonal $\Delta_X = \{(x,x) : x \in X\}$, which is a non-empty invariant set. Hence, $A = \bigcap_{n\in\mathbb{Z}}(T \times T)^n(W)$ is a non-empty, closed, invariant set since W is closed. Since W is open, it is an isolating neighborhood for A. If X were WM then $X \times X$ would be transitive and, being weakly rigid, $X \times X \times X \times X$ is pointwise recurrent and so $X \times X$ can contain no such isolated invariant set. □

Corollary 3.4. *Let (X,T) be a transitive cascade.*

(1) *If X is not connected, then its enveloping semigroup $E(X,T)$, as a dynamical system, is not WM.*
(2) *If X admits a proper, isolated, closed, invariant subset, then its enveloping semigroup $E(X,T)$, as a dynamical system, is not WM.*

Proof. By Corollary 3.2, (X,T) is weakly rigid and WM if $E(X,T)$ is WM. Now apply Theorem 3.3. □

Remark 3.5. We will next give an alternative proof of the fact that a weakly rigid WM system is necessarily connected; in fact we prove a slightly stronger result.

Theorem 3.6. *A totally transitive, weakly rigid cascade (X,T) is connected.*

Proof. Assume to the contrary that X is not connected. We first note that $\hat{X}$, the canonically defined largest totally disconnected factor of X, is nontrivial, and that $E(X) \to E(\hat{X})$. So we now assume that X is nontrivial and totally disconnected. Such a system always admits a nontrivial symbolic factor $X \to Y$ (i.e. $Y \subset \{0,1\}^{\mathbb{Z}}$ is a subshift), which by total transitivity is infinite. It then follows that there are four distinct points y_i $(i = 1,2,3,4)$ such that y_1 and y_2 are right asymptotic while y_3 and y_4 are left asymptotic (see [**GH**, Theorem 10.36], the so called Schwartzman lemma). Now clearly any limit point of the orbit $\{T^n(y_1,y_2,y_3,y_4) : n \in \mathbb{Z}\}$ has at most three distinct coordinates. In particular the point (y_1,y_2,y_3,y_4) is not recurrent. This implies that Y is not weakly rigid, contradicting the fact that Y is a factor of X. (See also Proposition 6.7 in [**GMa**].) □

4. The horocycle flow

THEOREM 4.1. *The enveloping semigroup of a classical horocycle flow* $(X, \{U_t\}_{t\in\mathbb{R}})$, *where* $G = PSL(2,\mathbb{R})$, $\Gamma < G$ *is a uniform lattice,* $X = G/\Gamma$ *and* $U_t(g\Gamma) = u_t g\Gamma = \left(\begin{smallmatrix} 1 & t \\ 0 & 1 \end{smallmatrix}\right) g\Gamma$, $t \in \mathbb{R}$, $g \in G$, *is WM. The same holds for the discrete flow* (X, U_1).

PROOF. Recall that $E(X, \{U_t\}_{t\in\mathbb{R}})$ is isomorphic, as a flow, to the infinite pointed product of the family of pointed systems $\{(X, x) : x \in X\}$, i.e. $E(X, \{U_t\}_{t\in\mathbb{R}}) = \bigvee_{x\in X}(X, x) = \overline{\{U_t\mathbf{x} : t \in \mathbb{R}\}} \subset X^X$, where $\mathbf{x} \in X^X$ is the identity map $\mathbf{x}(x) = x$, $\forall x \in X$. It follows that $E(X, \{U_t\}_{t\in\mathbb{R}})$ is the inverse limit of the family of finite pointed systems

$$\{X(\{x_1, x_2, \ldots, x_n\}) = (X, x_1) \vee (X, x_2) \vee \cdots \vee (X, x_n)\},$$

where we range over the directed collection of, unordered, k-tuples $\{x_1, x_2, \ldots, x_n\} \subset X$. It therefore suffices to show that each $X(\{x_1, x_2, \ldots, x_n\})$ is WM. The fact that this is indeed the case is a direct corollary of Ratner's theory, as follows:

By Ratner's orbit closure theorem (see [**R-91**, Theorem A]) we have

$$X(\{x_1, x_2, \ldots, x_n\}) = \overline{\{U_t X(x_1, x_2, \ldots, x_n) : t \in \mathbb{R}\}} = H(x_1, x_2, \ldots, x_n),$$

where $H < G \times G \times \cdots \times G$ (n times) is a closed connected subgroup of G^n containing the subgroup $\{u_t \times u_t \times \cdots \times u_t : t \in \mathbb{R}\}$ and there is a discrete uniform lattice $\Lambda < H$ so that $X(\{x_1, x_2, \ldots, x_n\}) = H(x_1, x_2, \ldots, x_n) \cong H/\Lambda$. By unique ergodicity of $(G/\Gamma, \{U_t\}_{t\in\mathbb{R}})$, the Haar measure λ on H/Λ is an n-fold self-joining of μ, the unique invariant probability measure on G/Γ. Now apply Ratner's joining theorem (see [**R-83**, Theorem 7, page 283]) to conclude that λ has the following form: There is, for some q, $1 \le q \le n$, a partition

$$I_j \subset I = \{1, 2, \ldots, n\},\ I_j = \{i_1^{(j)}, i_2^{(j)}, \ldots, i_{n_j}^{(j)}\},\ j = 1, \ldots q,$$

$$\bigcup_{j=1}^{q} I_j = I,$$

and there are elements $a_{i_k}^{(j)} \in G$, $k = 1, \ldots, n_j$, $j = 1, \ldots, q$, with

$$\Gamma_0^{(j)} = \bigcap_{k=1}^{n_j} a_{i_k}^{(j)} \Gamma (a_{i_k}^{(j)})^{-1}$$

uniform lattices, such that, each $(X^{I_j}, \lambda_j, \{U_t\}_{t\in\mathbb{R}})$ is isomorphic to the horocycle flow on $(G/\Gamma_0^{(j)}, \mu_0^{(j)}, \{U_t\}_{t\in\mathbb{R}})$, and $\lambda = \prod_{j=1}^q \lambda_j$. In particular then the measure preserving dynamical system $(H/\Lambda, \lambda, \{U_t\}_{t\in\mathbb{R}})$ is measure theoretically weakly mixing, hence also topologically WM.

Finally, the same arguments will work for the discrete flow (X, U_1). □

EXAMPLE 4.2. The case where Γ is a nonarithmetic maximal uniform lattice is special. We see that $E(X, \{U_t\}_{t\in\mathbb{R}}) = E(G/\Gamma, \{U_t\}_{t\in\mathbb{R}}) \cong X^{[X]}$, as dynamical systems, where $[X]$ denotes the collection of $\{U_t\}_{t\in\mathbb{R}}$ orbits in G/Γ. Note that both the $\mathbb{R}$-flow $E(X, \{U_t\}_{t\in\mathbb{R}})$ and the discrete system $E(X, U_1)$, have the property that any point $\mathbf{x} \in X^{[X]}$ whose coordinates have the property that no two of them belong to the same $\{U_t\}_{t\in\mathbb{R}}$ orbit, has a dense orbit in $X^{[X]}$. Of course card$([\mathrm{X}]) = \mathfrak{c}$, the cardinality of the continuum. Projecting on any set of four coordinates we obtain a real flow with 4-fold minimal self joinings. The corresponding U_1 cascade has the same property when we allow for off-diagonals resulting from the $\{U_t\}_{t\in\mathbb{R}}$ flow. This does not contradict the result of J. King [**K**, page 756], which says that there is no infinite minimal cascade (X,T) with 4-fold minimal self-joinings. The reason is that here, as it turns out, the 'future ε-bounded pair' produced in Theorem 21 of [**K**], must lie on the same $\{U_t\}_{t\in\mathbb{R}}$ orbit.

Remark 4.3. It is easy to see that if (X,T) is a dynamical system whose enveloping semigroup is WM then so is the associated system $(\hat{X},T)$ obtained from X by collapsing its mincenter to a point (see e.g. the remark at the beginning of Section 3). This leads us to the following:

Problem 4.4. Is there a nontrivial metric connected cascade (X,T) with trivial mincenter whose enveloping semigroup is WM ? Note that such a system is necessarily both proximal and weakly rigid.

References

[Ak-93] Ethan Akin, *The general topology of dynamical systems*, Graduate Studies in Mathematics, vol. 1, American Mathematical Society, Providence, RI, 1993. MR1219737

[Ak-96] Ethan Akin, *On chain continuity*, Discrete Contin. Dynam. Systems **2** (1996), no. 1, 111–120, DOI 10.3934/dcds.1996.2.111. MR1367390

[Ak-16] Ethan Akin, *Conjugacy in the Cantor set automorphism group*, Ergodic theory, dynamical systems, and the continuing influence of John C. Oxtoby, Contemp. Math., vol. 678, Amer. Math. Soc., Providence, RI, 2016, pp. 1–42. MR3589814

[AAG] Ethan Akin, Joseph Auslander, and Eli Glasner, *The topological dynamics of Ellis actions*, Mem. Amer. Math. Soc. **195** (2008), no. 913, vi+152, DOI 10.1090/memo/0913. MR2437846

[Aus] Joseph Auslander, *Minimal flows and their extensions*, North-Holland Mathematics Studies, vol. 153, North-Holland Publishing Co., Amsterdam, 1988. Notas de Matemática [Mathematical Notes], 122. MR956049

[AGW] J. Auslander, E. Glasner, and B. Weiss, *On recurrence in zero dimensional flows*, Forum Math. **19** (2007), no. 1, 107–114, DOI 10.1515/FORUM.2007.004. MR2296067

[DY] Tomasz Downarowicz and Xiangdong Ye, *When every point is either transitive or periodic*, Colloq. Math. **93** (2002), no. 1, 137–150, DOI 10.4064/cm93-1-9. MR1930259

[E] Robert Ellis, *Distal transformation groups*, Pacific J. Math. **8** (1958), 401–405. MR0101283

[F] Harry Furstenberg, *Disjointness in ergodic theory, minimal sets, and a problem in Diophantine approximation*, Math. Systems Theory **1** (1967), 1–49, DOI 10.1007/BF01692494. MR0213508

[G] E. Glasner, *Topological weak mixing and quasi-Bohr systems*, Israel J. Math. **148** (2005), 277–304, DOI 10.1007/BF02775440. Probability in mathematics. MR2191232

[GMa] S. Glasner and D. Maon, *Rigidity in topological dynamics*, Ergodic Theory Dynam. Systems **9** (1989), no. 2, 309–320, DOI 10.1017/S0143385700004983. MR1007412

[GMe] E. Glasner and M. Megrelishvili, *Hereditarily non-sensitive dynamical systems and linear representations*, Colloq. Math. **104** (2006), no. 2, 223–283, DOI 10.4064/cm104-2-5. MR2197078

[GH] Walter Helbig Gottschalk and Gustav Arnold Hedlund, *Topological dynamics*, American Mathematical Society Colloquium Publications, Vol. 36, American Mathematical Society, Providence, R. I., 1955. MR0074810

[K] Jonathan L. King, *A map with topological minimal self-joinings in the sense of del Junco*, Ergodic Theory Dynam. Systems **10** (1990), no. 4, 745–761, DOI 10.1017/S0143385700005873. MR1091424

[R-83] Marina Ratner, *Horocycle flows, joinings and rigidity of products*, Ann. of Math. (2) **118** (1983), no. 2, 277–313, DOI 10.2307/2007030. MR717825

[R-91] Marina Ratner, *Raghunathan's topological conjecture and distributions of unipotent flows*, Duke Math. J. **63** (1991), no. 1, 235–280, DOI 10.1215/S0012-7094-91-06311-8. MR1106945

[W] B. Weiss, *Multiple recurrence and doubly minimal systems*, Topological dynamics and applications (Minneapolis, MN, 1995), Contemp. Math., vol. 215, Amer. Math. Soc., Providence, RI, 1998, pp. 189–196, DOI 10.1090/conm/215/02940. MR1603185

The City College, 137 Street and Convent Avenue, New York City, New York 10031
Email address: ethanakin@earthlink.net

Department of Mathematics, Tel Aviv University, Tel Aviv, Israel
Email address: glasner@math.tau.ac.il

Institute of Mathematics, Hebrew University of Jerusalem, Jerusalem, Israel
Email address: weiss@math.huji.ac.il

Contemporary Mathematics
Volume **744**, 2020
https://doi.org/10.1090/conm/744/14928

The inhomogeneous Sprindžhuk conjecture over a local field of positive characteristic

Arijit Ganguly and Anish Ghosh

Abstract. We prove a strengthened version of the inhomogeneous Sprindžhuk conjecture in metric Diophantine approximation, over a local field of positive characteristic. The main tool is the transference principle of Beresnevich and Velani [**8**] coupled with earlier work of the second named author [**22**] who proved the standard, i.e. homogeneous version.

1. Introduction

The context of this paper is the metric theory of Diophantine approximation over local fields of positive characteristic. In [**22**], the second named author proved the Sprindžhuk conjectures in this setting (in fact, also in multiplicative form), here we prove the inhomogeneous variant of the conjecture. We use the inhomogeneous transference principle of Beresnevich and Velani [**8**] to transfer the homogeneous result from [**22**] and also use a positive characteristic version of the transference principle of Bugeaud and Laurent interpolating between uniform and standard Diophantine exponents, established recently by Bugeaud and Zhang [**10**]. The possibility of proving the S-arithmetic inhomogeneous Sprindžhuk conjectures was suggested by Beresnevich and Velani ([**8**], §8.4) and the present paper realises this expectation in another natural setting, that of local fields of positive characteristic.

Metric Diophantine approximation on manifolds is a subject which studies the extent to which typical Diophantine properties for Lebesgue measure on $\mathbb{R}^n$ are inherited by smooth submanifolds or other measures. The theory began with Mahler [**38**] who conjectured that almost every point on the Veronese curve is *not very well approximable*. Mahler's conjecture was resolved by Sprindžhuk [**42, 43**], who in turn made a stronger conjecture which was resolved by Kleinbock and Margulis [**33**] using methods from the ergodic theory of group actions on homogeneous spaces, specifically, sharp nondivergence estimates for unipotent flows on the space

2010 *Mathematics Subject Classification.* 11J83, 11K60, 37D40, 37A17, 22E40.
Key words and phrases. Diophantine approximation, dynamical systems.
The second named author gratefully acknowledges support from a grant from the Indo-French Centre for the Promotion of Advanced Research; a Department of Science and Technology, Government of India Swarnajayanti fellowship and a MATRICS grant from the Science and Engineering Research Board.

of lattices. Subsequently, an S-arithmetic version of the conjectures were established by Kleinbock and Tomanov [**35**] and a positive characteristic version was established by the second named author [**22**]. Both the latter works used adaptations of the dynamical approach of Kleinbock and Margulis. In [**8**], Beresnevich and Velani proved a transference principle which allowed them to prove an inhomogeneous versions of the Baker-Sprindžhuk conjectures. We refer the reader to the above papers for more details. We will recall all the relevant concepts in the function field context in the next section.

Following the work of Beresnevich and Velani, there have been several recent advances in inhomogeneous Diophantine approximation. In [**6**], an inhomogeneous Khintchine type theorem was established for affine subspaces, complementing the earlier work [**3**] for nondegenerate manifolds, see also [**27**] for more inhomogeneous results on affine subspaces. Further, an S-arithmetic inhomogeneous Khintchine type theorem for nondegenerate manifolds was established by Datta and the second named author [**12**].

1.1. The setup. We follow our paper [**17**] in setting the notation. Let p be a prime and $q := p^r$, where $r \in \mathbb{N}$, and consider the function field $\mathbb{F}_q(T)$. We define a function $|\cdot| : \mathbb{F}_q(T) \longrightarrow \mathbb{R}_{\geq 0}$ as follows.

$$|0| := 0 \text{ and } \left|\frac{P}{Q}\right| := e^{\deg P - \deg Q} \text{ for all nonzero } P, Q \in \mathbb{F}_q[T]\,.$$

Clearly $|\cdot|$ is a nontrivial, non-archimedian and discrete absolute value in $\mathbb{F}_q(T)$. This absolute value gives rise to a metric on $\mathbb{F}_q(T)$.

The completion field of $\mathbb{F}_q(T)$ is $\mathbb{F}_q((T^{-1}))$, i.e. the field of Laurent series over $\mathbb{F}_q$. The absolute value of $\mathbb{F}_q((T^{-1}))$, which we again denote by $|\cdot|$, is given as follows. Let $a \in \mathbb{F}_q((T^{-1}))$. For $a = 0$, define $|a| = 0$. If $a \neq 0$, then we can write

$$a = \sum_{k \leq k_0} a_k T^k \text{ where } k_0 \in \mathbb{Z},\, a_k \in \mathbb{F}_q \text{ and } a_{k_0} \neq 0\,.$$

We define k_0 as the *degree* of a, which will be denoted by $\deg a$, and $|a| := e^{\deg a}$. This clearly extends the absolute value $|\cdot|$ of $\mathbb{F}_q(T)$ to $\mathbb{F}_q((T^{-1}))$ and moreover, the extension remains non-archimedian and discrete. Let Λ and F denote $\mathbb{F}_q[T]$ and $\mathbb{F}_q((T^{-1}))$ respectively from now on. It is obvious that Λ is discrete in F. For any $d \in \mathbb{N}$, F^d is throughout assumed to be equipped with the supremum norm which is defined as follows

$$||\mathbf{x}|| := \max_{1 \leq i \leq n} |x_i| \text{ for all } \mathbf{x} = (x_1, x_2, ..., x_d) \in F^d\,,$$

and with the topology induced by this norm. Clearly Λ^n is discrete in F^n. Since the topology on F^n considered here is the usual product topology on F^n, it follows that F^n is locally compact as F is locally compact. Let λ be the Haar measure on F^d which takes the value 1 on the closed unit ball $||\mathbf{x}|| = 1$.

Diophantine approximation in the positive characteristic setting consists of approximating elements in F by 'rational' elements, i.e. those from $\mathbb{F}_q(T)$. This subject has been extensively studied, beginning with work of E. Artin [**1**] who developed the theory of continued fractions, and continuing with Mahler who developed Minkowski's geometry of numbers in function fields and Sprindžuk who, in addition to proving the analogue of Mahler's conjectures, also proved some transference principles in the function field setting (see [**42**]). The subject has also received considerable attention of late, we refer the reader to [**15**, **37**] for overviews and to [**2**, **18**, **28**, **29**, **36**] for a necessarily incomplete set of references.

In the present paper we prove an inhomogeneous analogue of the Sprindžuk conjectures, our main result is an upper bound for inhomogeneous Diophantine exponents.

THEOREM 1.1. *Let $U \subseteq F^d$ be open and $\mathbf{f} : U \longrightarrow F^n$ be a $(C, \alpha_0)-$good map, for some $C, \alpha_0 > 0$, and assume that $(\mathbf{f}, \lambda)$ is nonplanar. Then, for every $\theta \in F$, and λ almost every $\mathbf{x} \in U$,*

$$\omega(\mathbf{f}(\mathbf{x}), \theta) \leq 1.$$

We also establish the corresponding lower bound.

THEOREM 1.2. *Let $U \subseteq F^d$ be open and $\mathbf{f} : U \longrightarrow F^n$ be a $(C, \alpha_0)-$good map, for some $C, \alpha_0 > 0$, and assume that $(\mathbf{f}, \lambda)$ is nonplanar. Then, for every $\theta \in F$, and λ almost every $\mathbf{x} \in U$,*

$$\omega(\mathbf{f}(\mathbf{x}), \theta) \geq 1.$$

Remarks:

(1) Note that the exceptional set of $\mathbf{x}$ for which the inequalities in Theorems 1.1 and 1.2 need not hold depends on the inhomogeneous parameter θ.

(2) The relevant definitions are made in the next section. A main example to keep in mind is the original setup of Diophantine approximation on manifolds, i.e. if $\mathbf{f} = (f_1, \dots, f_n)$ where the f_i's are analytic and $1, f_1, \dots, f_n$ are linearly independent over F, then $\mathbf{f}$ is (C, α) good for some C, α and nonplanar. More generally, if $\mathbf{f}$ is a smooth nondegenerate map, then it is (C, α)-good as well as nonplanar. The notions of (C, α) good functions and nondegenerate maps were introduced by Kleinbock and Margulis [**33**].

(3) The homogeneous analogue of Theorem 1.1 was proved in [**22**] (Theorem 3.7), the lower bound is a consequence of Dirichlet's theorem.

(4) In [**3**], and subsequently in [**6**] a more general problem is considered where the inhomogeneous term is also allowed to vary. It should be possible to incorporate this improvement into Theorem 1.1.

(5) The next five sections deal with the proof of the main theorem. Sections 2 and 3 give the necessary prerequisites, in section 4 the lower bounds for Diophantine exponents are obtained and in section 6, the corresponding upper bounds. The final section is devoted to open questions and future possibilities for research.

Acknowledgments

Part of this work was done when both authors were visiting ICTS Bengaluru. We thank the institute for its hospitality and excellent working conditions. We thank the referee for many helpful suggestions which have improved the paper.

2. Homogeneous and Inhomogeneous Diophantine exponents

The theory of Diophantine approximation in positive characteristic begins with Dirichlet's theorem, which we now recall.

THEOREM 2.1. *(Theorem 2.1 [**17**]) Let $m, n \in \mathbb{N}$, $k = m + n$ and*

$$\mathfrak{a}^+ := \{\mathbf{t} := (t_1, t_2, \dots, t_k) \in \mathbb{Z}_+^k \, : \, \sum_{i=1}^{m} t_i = \sum_{j=1}^{n} t_{m+j}\} .$$

Consider m linear forms $Y_1, Y_2, \dots, Y_m$ over F in n variables. Then for any $\mathbf{t} \in \mathfrak{a}^+$, there exist solutions $\mathbf{q} = (q_1, q_2, \dots, q_n) \in \Lambda^n \setminus \{\mathbf{0}\}$ and $\mathbf{p} = (p_1, p_2, \dots, p_m) \in \Lambda^m$ of the following system of inequalities

$$\begin{cases} |Y_i\mathbf{q} - p_i| < e^{-t_i} & \text{for} \quad i = 1, 2, \dots, m \\ |q_j| \leq e^{t_{m+j}} & \text{for} \quad j = 1, 2, \dots, n \,. \end{cases} \tag{2.1}$$

We will consider only unweighted Diophantine approximation in this paper, so $t_1 = \cdots = t_m = 1/m$ and $t_{n+1} = \cdots = t_n = 1/n$. We denote by $\mathrm{M}_{m\times n}(F)$, the vector space of $m \times n$ matrices with entries from F equipped with the supremum norm. In view of Theorem 2.1, it is natural to define exponents of Diophantine approximation as follows. Let $X \in \mathrm{M}_{m\times n}(F)$ and $\boldsymbol{\theta} \in F^m$. The *inhomogeneous exponent*, $\omega(X, \boldsymbol{\theta})$ of X, is the supremum of the real numbers ω for which, for arbitrarily large $T \in \mathbb{N}$, the inequalities

$$\|X\mathbf{q} - \mathbf{p} - \boldsymbol{\theta}\| < e^{-\frac{n}{m}\omega T}, \qquad \|\mathbf{q}\| < e^T, \tag{2.2}$$

have a solution $(\mathbf{p}, \mathbf{q}) \in \Lambda^m \times (\Lambda^n \setminus \{\mathbf{0}\})$. The *uniform inhomogeneous exponent*, $\hat{\omega}(X, \boldsymbol{\theta})$, is the supremum of the real numbers $\hat{\omega}$ for which, for all sufficiently $T \in \mathbb{N}$, the inequalities

$$\|X\mathbf{q} - \mathbf{p} - \boldsymbol{\theta}\| < e^{-\frac{n}{m}\hat{\omega} T}, \qquad \|\mathbf{q}\| < e^T,$$

have a solution $(\mathbf{p}, \mathbf{q}) \in \Lambda^m \times (\Lambda^n \setminus \{\mathbf{0}\})$.

In this paper, we will adopt the point of view of Diophantine approximation of single linear forms, i.e. we will assume that $\mathbf{y} \in F^n$ where F^n is identified with $\mathrm{M}_{1\times n}(F)$ as opposed to simultaneous Diophantine approximation where one considers $\mathbf{y} \in \mathrm{M}_{n\times 1}$.

If $\theta = 0$, then the corresponding Diophantine exponent $\omega(\mathbf{y}) := \omega(\mathbf{y}, 0)$ (resp. $\hat{\omega}(\mathbf{y})$) is called the homogeneous Diophantine exponent. By Dirichlet's theorem stated above, $\omega(\mathbf{y}) \geq 1$ for every $\mathbf{y} \in F^n$. We are following the normalisation in [**8**] rather than the one used in [**22, 33**] according to which the critical exponent is n.

The Borel-Cantelli lemma implies that $\omega(\mathbf{y}) = 1$ for λ almost every $\mathbf{y} \in F^n$. It is therefore natural to define $\mathbf{y} \in F^n$ to be *very well approximable* if $\omega(\mathbf{y}) > 1$. Sprindžhuk [**43**] proved that for λ a.e. $x \in F$,

$$\mathbf{f}(x) := (x, x^2, \dots, x^n) \tag{2.3}$$

is not very well approximable, thereby settling the positive characteristic analogue of Mahler's conjecture. A special case of the theorems proved in this paper is that for every $\theta \in F$, $\omega(\mathbf{f}(x), \theta) = 1$ for almost every x. Following [**8**] we may define inhomogeneously extremal measures as follows.

DEFINITION 2.2. Let μ be a measure supported on a subset of F^n. We say that μ is inhomogeneously extremal if for all $\theta \in F$,

$$\omega(\mathbf{y}, \theta) = 1 \text{ for } \mu \text{ a.e. } \mathbf{y} \in F^n.$$

Then our main theorems can be restated as follows:

THEOREM 2.3. *Let $U \subseteq F^d$ be open and $\mathbf{f} : U \longrightarrow F^n$ be a (C, α_0)-good map, for some $C, \alpha_0 > 0$, and assume that $\mathbf{f}$ is nonplanar. Then $\mathbf{f}_*\lambda$ is inhomogeneously extremal.*

3. Good and nonplanar maps

We recall the following definitions and results from [**35**, §1 and 2]. For the sake of generality, we assume X is a Besicovitch metric space, $U \subseteq X$ is open, ν is a Radon measure on X, $(\mathcal{F}, |\cdot|)$ is a valued field and $f : X \longrightarrow \mathcal{F}$ is a given function such that $|f|$ is measurable. Recall that a metric space X is called *Besicovitch* [**35**] if there exists a constant N_X such that the following holds: for any bounded subset A of X and for any family $\mathcal{B}$ of nonempty open balls in X such that

$$\forall x \in A \text{ is a center of some ball of } B,$$

there is a finite or countable subfamily $\{B_i\}$ of B with

$$1_A \leq \sum_i 1_{B_i} \leq N_X.$$

For any $B \subseteq X$, we set

$$||f||_{\nu,B} := \sup_{x \in B \cap \text{ supp } (\nu)} |f(x)|.$$

DEFINITION 3.1. For $C, \alpha > 0$, f is said to be (C, α)-good on U with respect to ν if for every ball $B \subseteq U$ with center in supp (ν), one has

$$\nu(\{x \in B : |f(x)| < \varepsilon\}) \leq C \left(\frac{\varepsilon}{||f||_{\nu,B}} \right)^\alpha \nu(B)\,.$$

The following properties are immediate from Definition 3.1.

LEMMA 3.2. *Let $X, U, \nu, \mathcal{F}, f, C, \alpha,$ be as given above. Then one has*

(1) *f is (C, α)-good on U with respect to $\nu \Longleftrightarrow$ so is $|f|$.*
(2) *f is (C, α)-good on U with respect to $\nu \Longrightarrow$ so is cf for all $c \in \mathcal{F}$.*
(3) *$\forall\ i \in I, f_i$ are (C, α)-good on U with respect to ν and $\sup_{i \in I} |f_i|$ is measurable $\Longrightarrow$ so is $\sup_{i \in I} |f_i|$.*
(4) *f is (C, α)-good on U with respect to ν and $g : V \to \mathbb{R}$ is a continuous function such that $c_1 \leq |\frac{f}{g}| \leq c_2$ for some $c_1, c_2 > 0 \Longrightarrow g$ is $(C(\frac{c_2}{c_1})^\alpha, \alpha)$-good on U with respect to ν.*
(5) *Let $C_2 > 1$ and $\alpha_2 > 0$. Then f is (C_1, α_1)-good on U with respect to ν and $C_1 \leq C_2, \alpha_2 \leq \alpha_1 \Longrightarrow f$ is $(C_2, \alpha_2) -$ good on V with respect to ν.*

We say a map $\mathbf{f} = (f_1, f_2, \ldots, f_n)$ from U to $\mathcal{F}^n$, where $n \in \mathbb{N}$, is (C, α)-good on U with respect to ν, or simply $(\mathbf{f}, \nu)$ is (C, α)-good on U, if every $\mathcal{F}$-linear combination of $1, f_1, \ldots, f_n$ is (C, α)-good on U with respect to ν.

Definition 3.3. Let $\mathbf{f} = (f_1, f_2, \ldots, f_n)$ be a map from U to $\mathcal{F}^n$, where $n \in \mathbb{N}$. We say that $(\mathbf{f}, \nu)$ is *nonplanar* at a given point $x_0 \in U$ if for any ball B with centered at x_0, the restrictions of the functions $1, f_1, \ldots, f_n$ on $B \cap$ supp (ν) are linearly independent over $\mathcal{F}$. If $(\mathbf{f}, \nu)$ is nonplanar at ν almost every point of U, then it is called *nonplanar*. We also simply say $\mathbf{f}$ is *nonplanar* when there is no possibility of confusion.

A typical example is provided by $\mathbf{f} = (f_1, f_2, \ldots, f_n)$ where $1, f_1, \ldots, f_n$ are smooth and linearly independent on U. Such a map has been called *nondegenerate* by Kleinbock and Margulis.

For $m \in \mathbb{N}$ and a ball $B = B(x; r) \subseteq X$, where $x \in X$ and $r > 0$, we shall use the notation $3^m B$ to denote the ball $B(x; 3^m r)$. Finally, we will need the notion of a doubling measure.

Definition 3.4. The measure ν is said to be *doubling* on U if there exists $D > 0$ such that for every ball B with center in supp (ν) such that $2B \subseteq U$, one has
$$\frac{\nu(2B)}{\nu(B)} \leq D\,.$$

4. Transference principles and lower bounds

The lower bound will follow immediately from two Diophantine transference principles. The following result was proved by Bugeaud and Zhang [**10**] and constitutes a positive characteristic version of the transference principle of Bugeaud and Laurent [**9**].

Theorem 4.1. *(Theorem 1.2, [**10**]) Let $X \in F^{m \times n}$. Then for all $\boldsymbol{\theta} \in F^m$, we have*
$$\omega(X, \boldsymbol{\theta}) \geq \frac{1}{\hat{\omega}(X^t)} \text{ and } \hat{\omega}(X, \boldsymbol{\theta}) \geq \frac{1}{\omega(X^t)}. \tag{4.1}$$
with equalities for almost every $\boldsymbol{\theta}$.

We will also need a positive characteristic version of Dyson's transference principle [**14**] which can be formulated as follows.

Theorem 4.2. *For $\mathbf{y} \in F^n$,*
$$\omega(\mathbf{y}) = 1 \text{ if and only if } \omega({}^t\mathbf{y}) = 1.$$

We omit the short proof which can be obtained by a verbatim repetition of the proof in [**14**], or the more recent, more general version proved in Theorem 1.7 in [**11**].

It is now easy to complete the proof of the lower bound Theorem 1.2.

Proof. Under the hypothesis of Theorem 1.2, using Theorem 3.7 of [**22**], we have that for λ almost every $\mathbf{x}$, $\omega(\mathbf{f}(\mathbf{x})) = 1$. Set $y = \mathbf{f}(\mathbf{x})$, then by Dyson's transference principle, $\omega({}^t\mathbf{y}) = 1$. By Dirichlet's theorem, $\hat{\omega}(\mathbf{y}) \geq 1$ and the trivial inequality
$$\omega(\mathbf{y}, \theta) \geq \hat{\omega}(\mathbf{y}, \theta) \geq 0,$$
applied to ${}^t\mathbf{y}$ and $\theta = 0$ we get that $\omega({}^t\mathbf{y}) = 1$. Finally, by (4.1), we get that $\omega(\mathbf{y}, \theta) \geq 1$ which completes the proof.

□

5. The Transference principle of Beresnevich-Velani

In this section we state the inhomogeneous transference principle of Beresnevich and Velani from [**8**, Section 5] which will allow us to convert our inhomogeneous problem to the homogeneous one. Let (Ω, d) be a locally compact metric space. Given two countable indexing sets $\mathcal{A}$ and $\mathbf{T}$, let H and I be two maps from $\mathbf{T}\times\mathcal{A}\times\mathbb{R}_+$ into the set of open subsets of Ω such that

$$H \ : \ (t, \alpha, \varepsilon) \in \mathbf{T} \times \mathcal{A} \times \mathbb{R}_+ \to H_{\mathbf{t}}(\alpha, \varepsilon) \tag{5.1}$$

and

$$I \ : \ (t, \alpha, \varepsilon) \in \mathbf{T} \times \mathcal{A} \times \mathbb{R}_+ \to I_{\mathbf{t}}(\alpha, \varepsilon). \tag{5.2}$$

Furthermore, let

$$H_{\mathbf{t}}(\varepsilon) := \bigcup_{\alpha\in\mathcal{A}} H_{\mathbf{t}}(\alpha, \varepsilon) \text{ and } I_{\mathbf{t}}(\varepsilon) := \bigcup_{\alpha\in\mathcal{A}} I_{\mathbf{t}}(\alpha, \varepsilon). \tag{5.3}$$

Let Ψ denote a set of functions $\psi : \mathbf{T} \to \mathbb{R}_+ \ : \ \mathbf{t} \to \psi_{\mathbf{t}}$. For $\psi \in \Psi$, consider the limsup sets

$$\Lambda_H(\psi) = \limsup_{\mathbf{t}\in\mathbf{T}} H_{\mathbf{t}}(\psi_{\mathbf{t}}) \text{ and } \Lambda_I(\psi) = \limsup_{\mathbf{t}\in\mathbf{T}} I_{\mathbf{t}}(\psi_{\mathbf{t}}). \tag{5.4}$$

The sets associated with the map H will be called homogeneous sets and those associated with the map I, inhomogeneous sets. We now come to two important properties connecting these notions.

The intersection property. The triple (H, I, Ψ) is said to satisfy the intersection property if, for any $\psi \in \Psi$, there exists $\psi^* \in \Psi$ such that, for all but finitely many $\mathbf{t} \in \mathbf{T}$ and all distinct α and α' in $\mathcal{A}$, we have that

$$I_{\mathbf{t}}(\alpha, \psi_{\mathbf{t}}) \cap I_{\mathbf{t}}(\alpha', \psi_{\mathbf{t}}) \subset H_{\mathbf{t}}(\psi^*_{\mathbf{t}}). \tag{5.5}$$

The contraction property. Let μ be a finite, non atomic, doubling measure supported on a bounded subset $\mathbf{S}$ of Ω. We say that μ is contracting with respect to (I, Ψ) if, for any $\psi \in \Psi$, there exists $\psi^+ \in \Psi$ and a sequence of positive numbers $\{k_{\mathbf{t}}\}_{\mathbf{t}\in\mathbf{T}}$ satisfying

$$\sum_{\mathbf{t}\in\mathbf{T}} k_{\mathbf{t}} < \infty, \tag{5.6}$$

such that, for all but finitely $\mathbf{t} \in \mathbf{T}$ and all $\alpha \in \mathcal{A}$, there exists a collection $C_{\mathbf{t},\alpha}$ of balls B centred at $\mathbf{S}$ satisfying the following conditions:

$$\mathbf{S} \cap I_{\mathbf{t}}(\alpha, \psi_{\mathbf{t}}) \subset \bigcup_{B\in C_{\mathbf{t},\alpha}} B \tag{5.7}$$

$$\mathbf{S} \cap \bigcup_{B\in C_{\mathbf{t},\alpha}} B \subset I_{\mathbf{t}}(\alpha, \psi^+_{\mathbf{t}}) \tag{5.8}$$

and

$$\mu(5B \cap I_{\mathbf{t}}(\alpha, \psi_{\mathbf{t}})) \leq k_{\mathbf{t}} \mu(5B). \tag{5.9}$$

We are now in a position to state Theorem 5 from [**8**].

THEOREM 5.1. *Suppose that (H, I, Ψ) satisfies the intersection property and that μ is contracting with respect to (I, Ψ). Then*

$$\mu(\Lambda_H(\psi)) = 0 \ \forall\ \psi \in \Psi \Rightarrow \mu(\Lambda_I(\psi)) = 0 \ \forall\ \psi \in \Psi. \tag{5.10}$$

6. Proof of Theorem 1.1

Fix $\theta \in F$. It is enough to show that for any open ball $V \subseteq U$ such that $5V \subseteq U$, $\omega(\mathbf{f}(\mathbf{x}), \theta) \leq 1$ for λ almost all $\mathbf{x} \in V$. In fact, we prove

$$\forall\ \omega > 1, \lambda(\{\mathbf{x} \in V : \omega(\mathbf{f}(\mathbf{x}), \theta) > \omega\}) = 0.$$

For each $(t, \alpha = (p, \mathbf{q}), \varepsilon) \in \mathbb{N} \times (\Lambda \times \Lambda^n \setminus \{0\}) \times \mathbb{R}_+$, we set

$$I_{\mathbf{t}}(\alpha, \varepsilon) \stackrel{\text{def}}{=} \{\mathbf{x} \in V : |\mathbf{f}(\mathbf{x}) \cdot \mathbf{q} + p + \theta| < \varepsilon, ||\mathbf{q}|| \leq e^t\},$$

and

$$H_{\mathbf{t}}(\alpha, \varepsilon) \stackrel{\text{def}}{=} \{\mathbf{x} \in V : |\mathbf{f}(\mathbf{x}) \cdot \mathbf{q} + p| < \varepsilon, ||\mathbf{q}|| \leq e^t\}.$$

Let Ψ denote the collection of functions $\psi_\omega : \mathbb{N} \to \mathbb{R}, t \mapsto \frac{1}{e^{n\omega t}}$, for $\omega > 1$. We denote the restriction of λ to V by μ and thus it is supported on V.

Since $\forall\ \omega > 1, \{\mathbf{x} \in V : \omega(\mathbf{f}(\mathbf{x}), \theta) > \omega\}) \subseteq \Lambda_I(\psi_\omega)$ so, it suffices to show that $\lambda(\Lambda_I(\psi_\omega)) = 0$ for any $\omega > 1$. Theorem 3.7 in [**22**] implies that

$$\forall\ \omega > 1,\ \lambda(\Lambda_H(\psi_\omega)) = 0.$$

Therefore to prove Theorem 1.1, in view of the Theorem 5.1, we only need to verify the intersection and contraction properties. These will be performed in the following two subsections.

6.1. Verification of the intersection property. Let $t \in \mathbb{N}, \alpha = (p, \mathbf{q}), \alpha' = (p', \mathbf{q}') \in \Lambda \times \Lambda^n \setminus \{0\}$ with $\alpha \neq \alpha'$ and $\omega > 1$. If at least one of $||\mathbf{q}||$ and $||\mathbf{q}'||$ is $> e^t$, then there is nothing to prove. Otherwise, the ultrametric property yields that if $\mathbf{x} \in I_t(\alpha, \psi_\omega(t)) \cap I_t(\alpha', \psi_\omega(t))$ then

$$|\mathbf{f}(\mathbf{x}) \cdot (\mathbf{q} - \mathbf{q}') + (p - p')| \leq \max\{|\mathbf{f}(\mathbf{x}) \cdot \mathbf{q} + p + \theta|, |\mathbf{f}(\mathbf{x}) \cdot \mathbf{q}' + p' + \theta|\} \leq \tfrac{1}{e^{n\omega t}}. \tag{6.1}$$

Note that if $\mathbf{q} = \mathbf{q}'$, then $|p - p'| \leq \frac{1}{e^{n\omega t}}$ and so $p = p'$ which is impossible. Hence, it follows from (6.1) that $I_t(\alpha, \psi_\omega(t)) \cap I_t(\alpha', \psi_\omega(t)) \subseteq H_t(\alpha - \alpha', \psi_\omega(t))$.

6.2. Verification of the contraction property. Fix $\alpha \in \Lambda \times \Lambda^n \setminus \{0\}$. We observe that, for any $t \in \mathbb{N}$, $I_t(\alpha, \psi_\omega(t)) \subseteq I_t(\alpha, \psi_{\frac{\omega+1}{2}}(t))$ and

$$\mu(I_t(\alpha, \psi_{\frac{\omega+1}{2}}(t))) \leq \mu\left(\{\mathbf{x} \in V : |\mathbf{f}(\mathbf{x}) \cdot \mathbf{q} + p + \theta| < \frac{1}{e^{\frac{\omega+1}{2}nt}}\}\right) \ll \frac{1}{e^{\frac{\omega+1}{2}nt\alpha_0}} \mu(V), \tag{6.2}$$

since $\mathbf{f}$ is (C, α_0)-good on U. From the nonplanarity of $\mathbf{f}$, we have

$$\inf_{\alpha} \sup_{\mathbf{x} \in U} |\mathbf{f}(\mathbf{x}) \cdot \mathbf{q} + p + \theta| > 0.$$

So the absolute constant appearing in the last inequality of (6.2) can be made independent of α. Thus it turns out from (6.2) that, for all sufficiently large t,

$$I_t(\alpha, \psi_{\frac{\omega+1}{2}}(t)) \subsetneq V \text{ for all } \alpha. \tag{6.3}$$

For any t that satisfies (6.3) and all α, we now construct a collection of balls $C_{t,\alpha}$ centered in V which makes (5.7)-(5.9) hold. If $I_t(\alpha, \psi_\omega(t)) = \emptyset$ then we set $C_{t,\alpha}$ as the empty collection and consequently, (5.7)-(5.9) become trivial. Suppose $I_t(\alpha, \psi_\omega(t))$ is nonempty. Let $\mathbf{x} \in I_t(\alpha, \psi_\omega(t))$. Since $I_t(\alpha, \psi_{\frac{\omega+1}{2}}(t))$ is open, there exists a ball $B'(\mathbf{x})$ with center $\mathbf{x}$ such that $B'(\mathbf{x}) \subseteq I_t(\alpha, \psi_{\frac{\omega+1}{2}}(t))$. We can scale it and denote it by $B(\mathbf{x})$, due to (6.3), in such a way that

$$B(\mathbf{x}) \subseteq I_t(\alpha, \psi_{\frac{\omega+1}{2}}(t)) \not\supseteq V \cap 5B(\mathbf{x}). \tag{6.4}$$

It is also clear from the construction that $5B(\mathbf{x}) \subseteq 5V \subseteq U$. Consider

$$C_{t,\alpha} \stackrel{\text{def}}{=} \{B(\mathbf{x}) : \mathbf{x} \in I_t(\alpha, \psi_\omega(t))\}.$$

The conditions (5.7) and (5.8) are obvious.

Define $F_\alpha : U \longrightarrow \mathbb{R}, F_\alpha(\mathbf{x}) = |\mathbf{f}(\mathbf{x}) \cdot \mathbf{q} + p + \theta|,\ \forall\, \mathbf{x} \in U$ and let $B \in C_{t,\alpha}$. By the last inequality given in (6.4), we see that

$$\sup_{\mathbf{x}\in 5B} F_\alpha(\mathbf{x}) \geq \sup_{\mathbf{x}\in 5B\cap V} F_\alpha(\mathbf{x}) \geq \frac{1}{e^{\frac{\omega+1}{2}nt}}. \tag{6.5}$$

Furthermore, one has

$$\sup_{\mathbf{x}\in 5B\cap I_t(\alpha,\psi_\omega(t))} F_\alpha(\mathbf{x}) < \frac{1}{e^{n\omega t}} \leq \frac{1}{e^{n\omega t}} \times e^{\frac{\omega+1}{2}nt} \sup_{\mathbf{x}\in 5B} F_\alpha(\mathbf{x}) = \frac{1}{e^{\frac{w-1}{2}nt}} \sup_{\mathbf{x}\in 5B} F_\alpha(\mathbf{x}), \tag{6.6}$$

due to (6.5). Hence, from (6.6) and the assumption that $\mathbf{f}$ is (C, α_0)-good on U, it follows now that

$$\begin{aligned} \mu(5B \cap I_t(\alpha, \psi_\omega(t))) &= \lambda(5B \cap I_t(\alpha, \psi_\omega(t))) \\ &\leq \lambda\left(\{\mathbf{x} \in 5B : F_\alpha(\mathbf{x}) < \frac{1}{e^{\frac{w-1}{2}nt}} \sup_{\mathbf{x}\in 5B} F_\alpha(\mathbf{x})\}\right) \\ &\leq \frac{C}{e^{\frac{w-1}{2}nt\alpha_0}} \lambda(5B). \end{aligned} \tag{6.7}$$

Since $5B \cap V = V$ or $5B$, accordingly as $V \subseteq 5B$ or $5B \subseteq V$, so we have $\mu(5B) = \lambda(V)$ or $\lambda(5B)$. In the first case, we obtain $\lambda(5B) \leq \lambda(5V) = 5^n\lambda(V) = 5^n\mu(5B)$, and $\lambda(5B) = \mu(5B)$ in the later. Thus in either case, we see that $\lambda(5B) \leq 5^n\mu(5B)$. In view of this and (6.7), the condition (5.9) of the contraction property is obvious as soon as we set

$$k_t \stackrel{\text{def}}{=} \frac{5^n C}{e^{\frac{w-1}{2}nt\alpha_0}},\ \forall\, t \gg 1.$$

7. Further directions

In this section, we mention some directions for future research.

7.1. One vs almost every dichotomies. In [**32**], D. Kleinbock proved a remarkable *dichotomy* for Diophantine exponents. A special case of his results implies that if a connected analytic manifold $\mathcal{M} \subset \mathbb{R}^n$ has one not very well approximable point, then almost every point on $\mathcal{M}$ is not very well approximable. In [**13**], a p-adic version of this result was obtained. It is natural to ask if inhomogeneous analogues of Kleinbock's results hold. In other words, we propose

CONJECTURE 7.1. Let $\mathcal{M} \subset \mathbb{R}^n$ be a connected analytic manifold. Suppose there exists $\mathbf{x} \in \mathcal{M}$ such that for every $\theta \in \mathbb{R}$,

$$\omega(\mathbf{x}, \theta) = 1. \tag{7.1}$$

Then $\mathcal{M}$ is inhomogeneously extremal.

This conjecture can of course be formulated over any local field as well as in the multiplicative setting. It should be noted that Kleinbock's technique does not seem to be directly applicable in the inhomogeneous setting.

7.2. Diophantine approximation on limit sets. Beginning with pioneering work of Patterson [**39**], the theory of metric Diophantine approximation in the context of dense orbits of geometrically finite Kleinian groups has developed into a full fledged theory. Recently, in [**7**], a theory of metric Diophantine approximation on manifolds was developed in the context of Kleinian groups. Namely, questions of inheritance of Diophantine properties for proper subsets of the limit set of a Kleinian group were investigated. This theory has a natural counterpart in positive characteristic; where one considers orbits of discrete subgroups of $G(k)$ for algebraic groups G defined over k on the boundary of the Bruhat-Tits building. It would be interesting to obtain a "manifold" theory in this context analogous to [**7**].

7.3. Friendly and nonplanar measures and multiplicative Diophantine approximation. It should be possible to extend our main Theorem to a wider class of measures, namely strongly contracting measures as considered by Beresnevich and Velani [**8**]. This class of measures includes friendly measures as defined by Kleinbock, Lindenstrauss and Weiss [**34**]. Though we do not discuss this here, in fact Theorems 1.1 and 1.2 should hold for a wider class of measures, the so called *strongly contracting measures* as introduced by Beresnevich and Velani, a category which includes the important class of *friendly* measures introduced earlier by Kleinbock, Lindenstrauss and Weiss [**34**]. It should also be possible to extend the main Theorem to the setting of multiplicative Diophantine approximation, thereby obtaining an inhomogeneous analogue of Baker's strong extremality conjecture.

7.4. Khintchine-Groshev type theorems. In [**36**], S. Kristensen proves an asymptotic formula for the number of solutions to inhomogeneous Khintchine type inequalities for matrices with entries in F, thereby obtaining an analogue of W. Schmidt's results [**40, 41**] in the positive characteristic setting. While this generality seems out of reach at present in the context of manifolds, it would be interesting to prove a qualitative result, namely homogeneous and inhomogeneous Khintchine type theorems for smooth manifolds in the positive characteristic setting. These would constitute function field analogues of the work of Bernik, Kleinbock and Margulis [**5**] who proved the convergence Khintchine theorem for smooth nondegenerate manifolds, and Beresnevich, Bernik, Kleinbock and Margulis [**4**] who proved the divergence case. In the inhomogeneous case, the convergence and divergence khintchine type theorems were proved by Badziahin, Beresnevich and Velani [**3**].

7.5. Affine subspaces and their nondegenerate submanifolds. The results in the present paper have to do with nondegenerate manifolds. At the other end of the spectrum lie affine subspaces, the study of whose Diophantine properties involves subtle considerations concerning the slope of the subspace. There has been considerable work in this area recently, cf. [**20**, **21**, **23–25**, **30**, **31**]. We refer the reader to [**26**] for a survey of this subject. It would be interesting to obtain function field analogues of these results, both homogeneous and inhomogeneous.

References

[1] E. Artin, *Quadratische Körper im Gebiete der höheren Kongruenzen. I* (German), Math. Z. **19** (1924), no. 1, 153–206, DOI 10.1007/BF01181074. MR1544651

[2] J. S. Athreya, Anish Ghosh, and Amritanshu Prasad, *Ultrametric logarithm laws, II*, Monatsh. Math. **167** (2012), no. 3-4, 333–356, DOI 10.1007/s00605-012-0376-y. MR2961287

[3] Dzmitry Badziahin, Victor Beresnevich, and Sanju Velani, *Inhomogeneous theory of dual Diophantine approximation on manifolds*, Adv. Math. **232** (2013), 1–35, DOI 10.1016/j.aim.2012.09.022. MR2989975

[4] V. V. Beresnevich, V. I. Bernik, D. Y. Kleinbock, and G. A. Margulis, *Metric Diophantine approximation: the Khintchine-Groshev theorem for nondegenerate manifolds*, Mosc. Math. J. **2** (2002), no. 2, 203–225, DOI 10.17323/1609-4514-2002-2-2-203-225. Dedicated to Yuri I. Manin on the occasion of his 65th birthday. MR1944505

[5] V. Bernik, D. Kleinbock, and G. A. Margulis, *Khintchine-type theorems on manifolds: the convergence case for standard and multiplicative versions*, Internat. Math. Res. Notices **9** (2001), 453–486, DOI 10.1155/S1073792801000241. MR1829381

[6] V. Beresnevich, A. Ganguly, A. Ghosh and S. Velani, *Inhomogeneous Dual Diophantine Approximation on Affine Subspaces*, Int. Math. Res. Not., doi:10.1093/imrn/rny124, 32 pp.

[7] Victor Beresnevich, Anish Ghosh, David Simmons, and Sanju Velani, *Diophantine approximation in Kleinian groups: singular, extremal, and bad limit points*, J. Lond. Math. Soc. (2) **98** (2018), no. 2, 306–328, DOI 10.1112/jlms.12133. MR3873110

[8] Victor Beresnevich and Sanju Velani, *An inhomogeneous transference principle and Diophantine approximation*, Proc. Lond. Math. Soc. (3) **101** (2010), no. 3, 821–851, DOI 10.1112/plms/pdq002. MR2734962

[9] Yann Bugeaud and Michel Laurent, *On exponents of homogeneous and inhomogeneous Diophantine approximation*, Mosc. Math. J. **5** (2005), no. 4, 747–766, 972, DOI 10.17323/1609-4514-2005-5-4-747-766. MR2266457

[10] Y. Bugeaud and Z. Zhang, *On homogeneous and inhomogeneous Diophantine approximation over the fields of formal power series*, https://arxiv.org/abs/1902.09034.

[11] S. Chow, A. Ghosh, L. Guan, A. Marnat and D. Simmons, *Diophantine transference inequalities: weighted, inhomogeneous, and intermediate exponents*, to appear in Annali della Scuola Normale Superiore di Pisa, https://arxiv.org/abs/1808.07184.

[12] S. Datta and A. Ghosh, *S-arithmetic inhomogeneous Diophantine approximation on manifolds*, https://arxiv.org/abs/1801.08848.

[13] S. Datta and A. Ghosh, *Diophantine inheritance for p-adic measures*, https://arxiv.org/abs/1903.09362.

[14] F. J. Dyson, *On simultaneous Diophantine approximations*, Proc. London Math. Soc. (2) **49** (1947), 409–420, DOI 10.1112/plms/s2-49.6.409. MR0022868

[15] Bernard de Mathan, *Approximations diophantiennes dans un corps local* (French), Bull. Soc. Math. France Suppl. Mém. **21** (1970), 93. MR0274396

[16] Lior Fishman, David Simmons, and Mariusz Urbański, *Diophantine approximation and the geometry of limit sets in Gromov hyperbolic metric spaces*, Mem. Amer. Math. Soc. **254** (2018), no. 1215, v+137, DOI 10.1090/memo/1215. MR3826896

[17] Arijit Ganguly and Anish Ghosh, *Dirichlet's theorem in function fields*, Canad. J. Math. **69** (2017), no. 3, 532–547, DOI 10.4153/CJM-2016-024-2. MR3679686

[18] Arijit Ganguly and Anish Ghosh, *Remarks on Diophantine approximation in function fields*, Math. Scand. **124** (2019), no. 1, 5–14, DOI 10.7146/math.scand.a-109985. MR3904745

[19] Arijit Ganguly and Anish Ghosh, *Quantitative Diophantine approximation on affine subspaces*, Math. Z. **292** (2019), no. 3-4, 923–935, DOI 10.1007/s00209-018-2115-0. MR3980277

[20] Anish Ghosh, *A Khintchine-type theorem for hyperplanes*, J. London Math. Soc. (2) **72** (2005), no. 2, 293–304, DOI 10.1112/S0024610705006587. MR2156655

[21] Anish Ghosh, *Dynamics of homogeneous spaces and Diophantine approximation on manifolds*, ProQuest LLC, Ann Arbor, MI, 2006. Thesis (Ph.D.)–Brandeis University. MR2708455

[22] Anish Ghosh, *Metric Diophantine approximation over a local field of positive characteristic*, J. Number Theory **124** (2007), no. 2, 454–469, DOI 10.1016/j.jnt.2006.10.009. MR2321374

[23] Anish Ghosh, *A Khintchine-Groshev theorem for affine hyperplanes*, Int. J. Number Theory **7** (2011), no. 4, 1045–1064, DOI 10.1142/S1793042111004228. MR2812651

[24] Anish Ghosh, *Diophantine approximation on affine hyperplanes*, Acta Arith. **144** (2010), no. 2, 167–182, DOI 10.4064/aa144-2-6. MR2669718

[25] Anish Ghosh, *Diophantine exponents and the Khintchine Groshev theorem*, Monatsh. Math. **163** (2011), no. 3, 281–299, DOI 10.1007/s00605-010-0239-3. MR2805874

[26] Anish Ghosh, *Diophantine approximation on subspaces of $\mathbb{R}^n$ and dynamics on homogeneous spaces*, Handbook of group actions. Vol. IV, Adv. Lect. Math. (ALM), vol. 41, Int. Press, Somerville, MA, 2018, pp. 509–527. MR3888693

[27] Anish Ghosh and Antoine Marnat, *On Diophantine transference principles*, Math. Proc. Cambridge Philos. Soc. **166** (2019), no. 3, 415–431, DOI 10.1017/S0305004118000014. MR3933905

[28] Anish Ghosh and Robert Royals, *An extension of the Khinchin-Groshev theorem*, Acta Arith. **167** (2015), no. 1, 1–17, DOI 10.4064/aa167-1-1. MR3310488

[29] Dong Han Kim and Hitoshi Nakada, *Metric inhomogeneous Diophantine approximation on the field of formal Laurent series*, Acta Arith. **150** (2011), no. 2, 129–142, DOI 10.4064/aa150-2-3. MR2836382

[30] D. Kleinbock, *Extremal subspaces and their submanifolds*, Geom. Funct. Anal. **13** (2003), no. 2, 437–466, DOI 10.1007/s000390300011. MR1982150

[31] Dmitry Kleinbock, *An extension of quantitative nondivergence and applications to Diophantine exponents*, Trans. Amer. Math. Soc. **360** (2008), no. 12, 6497–6523, DOI 10.1090/S0002-9947-08-04592-3. MR2434296

[32] Dmitry Kleinbock, *An 'almost all versus no' dichotomy in homogeneous dynamics and Diophantine approximation*, Geom. Dedicata **149** (2010), 205–218, DOI 10.1007/s10711-010-9477-8. MR2737689

[33] D. Y. Kleinbock and G. A. Margulis, *Flows on homogeneous spaces and Diophantine approximation on manifolds*, Ann. of Math. (2) **148** (1998), no. 1, 339–360, DOI 10.2307/120997. MR1652916

[34] Dmitry Kleinbock, Elon Lindenstrauss, and Barak Weiss, *On fractal measures and Diophantine approximation*, Selecta Math. (N.S.) **10** (2004), no. 4, 479–523, DOI 10.1007/s00029-004-0378-2. MR2134453

[35] Dmitry Kleinbock and George Tomanov, *Flows on S-arithmetic homogeneous spaces and applications to metric Diophantine approximation*, Comment. Math. Helv. **82** (2007), no. 3, 519–581, DOI 10.4171/CMH/102. MR2314053

[36] S. Kristensen, *Metric inhomogeneous Diophantine approximation in positive characteristic*, Math. Scand. **108** (2011), no. 1, 55–76, DOI 10.7146/math.scand.a-15160. MR2780807

[37] A. Lasjaunias, *A survey of Diophantine approximation in fields of power series*, Monatsh. Math. **130** (2000), no. 3, 211–229, DOI 10.1007/s006050070036. MR1777096

[38] Kurt Mahler, *Über das Maßder Menge aller S-Zahlen* (German), Math. Ann. **106** (1932), no. 1, 131–139, DOI 10.1007/BF01455882. MR1512754

[39] S. J. Patterson, *Diophantine approximation in Fuchsian groups*, Philos. Trans. Roy. Soc. London Ser. A **282** (1976), no. 1309, 527–563, DOI 10.1098/rsta.1976.0063. MR0568140

[40] Wolfgang Schmidt, *A metrical theorem in diophantine approximation*, Canadian J. Math. **12** (1960), 619–631, DOI 10.4153/CJM-1960-056-0. MR118711

[41] Wolfgang M. Schmidt, *Metrical theorems on fractional parts of sequences*, Trans. Amer. Math. Soc. **110** (1964), 493–518, DOI 10.2307/1993695. MR159802

[42] V. G. Sprindžuk, *Mahler's problem in metric number theory*, Translated from the Russian by B. Volkmann. Translations of Mathematical Monographs, Vol. 25, American Mathematical Society, Providence, R.I., 1969. MR0245527

[43] V. G. Sprindžuk, *Achievements and problems of the theory of Diophantine approximations* (Russian), Uspekhi Mat. Nauk **35** (1980), no. 4(214), 3–68, 248. MR586190

Department of Mathematics, IME Building Indian Institute of Technology Kanpur Kanpur, P.O.- IIT Kanpur, P.S.- Kalyanpur District - Kanpur Nagar Pin - 208016 Uttar Pradesh, India

Email address: arijit.ganguly1@gmail.com

School of Mathematics, Tata Institute of Fundamental Research, Homi Bhabha Road, Navy Nagar, Colaba, Mumbai 400005, India

Email address: ghosh@math.tifr.res.in

Contemporary Mathematics
Volume **744**, 2020
https://doi.org/10.1090/conm/744/14986

Dynamical generation of parameter laminations

Alexander Blokh, Lex Oversteegen, and Vladlen Timorin

Dedicated to the memory of Sergiy Kolyada

Abstract. Local similarity between the Mandelbrot set and quadratic Julia sets manifests itself in a variety of ways. We discuss a combinatorial one, in the language of geodesic laminations. More precisely, we compare quadratic invariant laminations representing Julia sets with the so-called Quadratic Minor Lamination (QML) representing a locally connected model of the Mandelbrot set. Similarly to the construction of an invariant lamination by pullbacks of certain leaves, we describe how QML can be generated by properly understood pullbacks of certain minors. In particular, we show that the minors of all non-renormalizable quadratic laminations can be obtained by taking limits of "pullbacks" of minors from the main cardioid.

Introduction

Quadratic polynomials $P_c(z) = z^2 + c$, where $c \in \mathbb{C}$, play an important role in complex dynamics. They provide a simple but highly non-trivial example of polynomial dynamical systems (note that every quadratic polynomial is affinely conjugate to one of the form P_c), and this family is universal in the sense that many properties of the c-parameter plane reappear locally in almost any analytic family of holomorphic maps [**McM00**]. The central object in the c-plane is the *Mandelbrot set* $\mathcal{M}_2$. By definition, $c \in \mathcal{M}_2$ if the Julia set $J(P_c)$ of P_c is connected, equivalently, if the sequence of iterates $P_c^n(c)$ does not escape to infinity (see [**DH85**]).

The Mandelbrot set is compact and connected. It is not known if it is locally connected, but there is a nice model $\mathcal{M}_2^c$, due to Douady, Hubbard and Thurston, of $\mathcal{M}_2$ (i.e., there exists a continuous map $\pi : \mathcal{M}_2 \to \mathcal{M}_2^c$ such that point inverses are connected); moreover, if $\mathcal{M}_2$ is locally connected, π is a homeomorphism. Namely, set $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$ and $\mathbb{S} = \{z \in \mathbb{C} \,|\, |z| = 1\}$; call $\mathbb{D}$ the *unit disk* and $\mathbb{S}$ the *unit circle*. There are pairwise disjoint chords (including degenerate chords, i.e. singletons in $\mathbb{S}$) or polygons inscribed in $\overline{\mathbb{D}} = \{z \in \mathbb{C} : |z| \leqslant 1\}$ such that, after collapsing all these chords and polygons to points, we get a quotient space $\mathcal{M}_2^c$. We will write QML for the set consisting of all these chords and edges of all these polygons. This set is called the *quadratic minor lamination*.

2010 *Mathematics Subject Classification.* Primary 37F20; Secondary 37F10, 37F50.

Key words and phrases. Complex dynamics; laminations; Mandelbrot set; Julia set.

The second named author was partially supported by NSF grant DMS–1807558.

The third named author was partially supported within the framework of the HSE University Basic Research Program and the Russian Academic Excellence Project '5-100'.

More generally, a (geodesic) *lamination* is a set of chords (called *leaves*) in $\overline{\mathbb{D}}$ that contains all points of $\mathbb{S}$ such that the limit of any converging sequence of leaves is a leaf. The lamination QML can be described explicitly. For example, one can algorithmically generate countably many leaves dense in QML, and there are several known constructions, e.g. [**Lav86**, **Lav89**] (other combinatorial viewpoints on $\mathcal{M}_2^c$ and QML can be found in [**BOPT16**, **Kel00**, **PR08**, **Sch09**]). In this paper, a new construction is provided that is based on taking preimages under the angle doubling map. Each of the sets $\mathcal{M}_2$ and $\mathcal{M}_2^c$ contains countable and dense family of homeomorphic copies of itself. Thus, $\mathcal{M}_2$ and $\mathcal{M}_2^c$ are examples of so-called *fractal* sets.

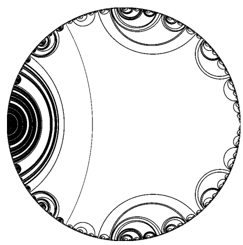


FIGURE 1. The geolamination QML

A description of QML by Thurston [**Thu85**] refers to laminational models of Julia sets. By the *filled Julia set* $K(P_c)$ of a polynomial P_c we mean the set of points $z \in \mathbb{C}$ with $P_c^n(z) \not\to \infty$. The *Julia set* $J(P_c)$ is the boundary of $K(P_c)$. If $K(P_c)$ is locally connected, then it can be also obtained from $\overline{\mathbb{D}}$ by collapsing leaves and finite polygons of some lamination $\mathcal{L}(P_c)$.

Indeed, if $K(P_c)$ is locally connected, the Riemann map defined for the complement of $K(P_c)$ can be extended onto $\mathbb{S}$ which gives rise to a continuous map $\psi : \mathbb{S} \to J(P_c)$ that semiconjugates the angle doubling map $\sigma_2 : \mathbb{S} \to \mathbb{S}$ (taking $z \in \mathbb{S}$ to z^2) and $P_c|_{J(P_c)}$. Considering convex hulls of *fibers* (point-inverses) of ψ and collecting boundary edges of these convex hulls, we obtain the lamination $\mathcal{L}(P_c)$. Declaring points x, y of $\mathbb{S}$ *equivalent* if and only if $\psi(x) = \psi(y)$ we arrive at the *invariant laminational equivalence* $\sim_c$ and the associated quotient space $J_{\sim_c}$ of $\mathbb{S}$ (the *topological Julia set*), homeomorphic to $J(P_c)$. Equivalence classes of $\sim_c$ have pairwise disjoint convex hulls. The *topological polynomial* $f_{\sim_c} : J_{\sim_c} \to J_{\sim_c}$, induced by σ_2, is topologically conjugate to $P_c|_{J(P_c)}$. Laminational equivalence relations $\sim$ similar to $\sim_c$ can be introduced with no references to polynomials by listing their properties similar to those of $\sim_c$ (this can be done for degrees higher than 2 as well). In that case one also considers the collection $\mathcal{L}_\sim$ of the edges of convex hulls of all $\sim$-classes and all singletons in $\mathbb{S}$ called the *q-lamination (generated by $\sim$)*.

A lamination $\mathcal{L}_{\sim_c}$ thus obtained satisfies certain dynamical properties (in our presentation we rely upon [**BMOV13**]). Below we think of σ_2 applied to a chord ℓ with endpoints a and b so that it maps to the chord whose endpoints are $\sigma_2(a)$ and $\sigma_2(b)$; we can think of this as an extension of σ_2 over ℓ and make it linear on ℓ. The properties are as follows:

(1) **forward invariance:** for every $\ell \in \mathcal{L}$, we have $\sigma_2(\ell) \in \mathcal{L}$;

(2) **backward invariance:** for every $\ell \in \mathcal{L}$ we have $\ell = \sigma_2(\ell_1)$ for some $\ell_1 \in \mathcal{L}$;
(3) **sibling property:** for every $\ell \in \mathcal{L}$, we have $-\ell \in \mathcal{L}$.

Here $-\ell$ is the image of ℓ under the map $z \mapsto -z$ of $\mathbb{S}$. (Under this map all angles are incremented by $\frac{1}{2}$ modulo 1). The leaf $-\ell$ is called the *sibling* of ℓ. A chord which is a diameter of $\mathbb{D}$ is said to be *critical.* Laminations with properties (1)–(3) are called *quadratic invariant laminations.* By [**BMOV13**] all quadratic q-laminations $\mathcal{L}_\sim$ are invariant, however the converse is not true and there are quadratic invariant laminations that are not q-laminations. Below we often call quadratic invariant laminations simply *quadratic laminations.*

Properties (1) – (3) from above deal exclusively with leaves. To understand the dynamics one also considers components of the complement in $\overline{\mathbb{D}}$ to the union of all leaves of $\mathcal{L}$. More precisely, a *gap* of $\mathcal{L}$ is the closure of a component of $\mathbb{D} \setminus \bigcup_{\ell \in \mathcal{L}} \ell$. Gaps G are said to be *finite* or *infinite* according to whether $G \cap \mathbb{S}$ is a finite or infinite set. By [**BMOV13**] if G is a gap of a quadratic lamination $\mathcal{L}$, then either all its edges map to one leaf of $\mathcal{L}$, or all its edges map to a single point in $\mathbb{S}$, or the convex hull of the set $\sigma_2(G\cap\mathbb{S})$ is a gap of $\mathcal{L}$ which one can view as the *image* of G. Moreover, the map on the boundary of G satisfies **gap invariance**: either there exists a critical edge of G, or the map $\tau = \sigma_2|_{G\cap\mathbb{S}}$ extends to $\mathbb{S}$ as an orientation preserving covering map $\hat{\tau}$ such that $G\cap\mathbb{S}$ is the full preimage of $\tau(G\cap\mathbb{S})$ under $\hat{\tau}$. Gap invariance was part of the original definition of a (geodesic) lamination given by Thurston in [**Thu85**]. It allows us to extend the map σ_2 onto the entire $\overline{\mathbb{D}}$ if a quadratic lamination $\mathcal{L}$ is given. Indeed, we have already described how σ_2 acts on leaves; it can then be extended over gaps using the *barycentric* construction (see [**Thu85**] for details).

Due to the backward invariance property, quadratic laminations can often be generated by taking pullbacks of leaves. By a *pullback* of a leaf $\ell \in \mathcal{L}$, we mean a leaf $\ell_1 \in \mathcal{L}$ such that $\sigma_2(\ell_1) = \ell$. An *iterated pullback* of ℓ of level n is defined as a leaf $\ell_n \in \mathcal{L}$ with $\sigma_2^n(\ell_n) = \ell$. The concept of (iterated) pullback is widely used in the study of (quadratic) invariant laminations. In this paper we show that it can also be used as one studies *parameter* laminations, i.e., laminations which do not satisfy conditions (1) — (3), such as QML. Let us now discuss QML in more detail.

To measure arc lengths on $\mathbb{S}$, we use the normalized Lebesgue measure (the total length of $\mathbb{S}$ is 1). The *length of a chord* is by definition the length of the shorter circle arc connecting its endpoints. Following Thurston, define a *major leaf* (a *major*) of a quadratic lamination as a longest leaf of it. (There may be one longest leaf that is critical or two longest leaves that are siblings.) The *minor leaf* (the *minor*) of a lamination is the σ_2-image of a major. If a minor m is non-periodic, then there exists a unique maximal lamination with minor m denoted by $\mathcal{L}(m)$. If a minor m is periodic and non-degenerate, then we define $\mathcal{L}(m)$ as the unique *q-lamination* with minor m. Finally, if m is a periodic singleton, then we explicitly define $\mathcal{L}(m)$ later in the paper so that m is the minor of $\mathcal{L}(m)$ (note, that in this case the choice of $\mathcal{L}(m)$ is irrelevant for our purposes). Call $\mathcal{L}(m)$ the *minor leaf lamination associated with m.* Observe that there are no minors that are non-degenerate and have exactly one periodic endpoint.

A chord in $\overline{\mathbb{D}}$ with endpoints a and b is denoted by ab. If two distinct chords intersect in $\mathbb{D}$, we say that they *cross* or that they are *linked.* Given a chord ab, without a lamination, we have ambiguity in defining pullbacks of ab. Namely, there are two preimages of a and two preimages of b, and, in general, there are several ways of connecting the preimages of a with the preimages of b. Even if we prohibit crossings and impose the sibling property, then there are three ways (two ways of connecting the preimages by two chords and one way of connecting them by four chords). However, if we know that the pullbacks must belong to $\mathcal{L}(m)$, then they are well defined. We can describe the process of taking pullbacks explicitly, without referring to $\mathcal{L}(m)$. One of the main objectives of this paper is to apply a similar pullback construction to QML.

Thurston's definition of QML is simply the following: QML consists precisely of the minors of all quadratic laminations. In particular, it is true (although not at all obvious) that different minors do not cross.

Offsprings of a minor. In order to state the first main result, we introduce some terminology and notation. The convex hull of a subset $A \subset \mathbb{R}^2 = \mathbb{C}$ will be denoted by $\mathrm{CH}(A)$. Let ℓ and ℓ_1 be chords of $\mathbb{S}$, possibly degenerate, not passing through the center of the disk. We will write $H(\ell)$ for the smaller open circle arc bounded by the endpoints of ℓ. Set $D(\ell) = \mathrm{CH}(H(\ell))$; since $H(\ell)$ is an open arc, $D(\ell)$ does not include ℓ. If $\ell_1 \in D(\ell)$, then we write $\ell_1 < \ell$. The notation $\ell_1 \leqslant \ell$ will mean $\ell_1 \subset \overline{D(\ell)}$. Note that, if ℓ_1 shares just one endpoint with ℓ and $\ell_1 \leqslant \ell$, then it is not true that $\ell_1 < \ell$. It follows that if $\ell_1 \leqslant \ell, \ell_1 \neq \ell$ then $|\ell_1| < |\ell|$, where $|\ell|$ denotes the length of ℓ; in particular $\ell_1 < \ell$ implies $|\ell_1| < |\ell|$. If $\ell_1 < \ell$ (resp., $\ell_1 \leqslant \ell$), then we say that ℓ_1 lies *strictly behind* (resp., *behind*) ℓ. Observe that our terminology applies to degenerate chords (i.e., singletons in the unit circle) too; a degenerate chord $\ell_1 = \{b\}$ is strictly behind ℓ if and only if $b \in H(\ell)$, and $\ell_1 \leqslant \ell$ simply means that $b \in \overline{H(\ell)}$.

Let us now describe an inductive process that shows how *dynamical* pullbacks of minors of quadratic laminations lead to the construction of the *parametric* lamination QML. Namely, consider any non-degenerate minor $m \in$ QML. Suppose that a point $a \in \mathbb{S}$ lies behind m and $\sigma_2^n(a)$ is an endpoint of m for some minimal $n > 0$. Observe that then a is not periodic as no image of a minor is located behind this minor. Consider *all* numbers k such that $\sigma_2^k(a)$ is an endpoint of a minor m'_k with $a < m'_k \leqslant m$ (thus, a is separated from m by m'_k or $m'_k = m$), and the least such number l. Denote by m_a the pullback of m'_l in $\mathcal{L}(m'_l)$ containing a such that $\sigma_2^{l-1}(m_a)$ is a major of $\mathcal{L}(m'_l)$ and call it an *offspring* of m. We also say that m_a is a *child* of m'_l. Observe that periodic minors are nobody's offsprings. Indeed, if $m' \leqslant m'', m' \neq m''$ are minors, $\sigma_2^i(m') = m''$, and m' is periodic, then $\sigma_2^j(m'') = m' \leqslant m''$ for some j, and it is well-known that this is impossible for minors.

Theorem A. Let $m \in$ QML be a non-degenerate minor. Then offsprings of a minor $m \in$ QML are minors too (i.e., they are leaves of QML). Thus, if a point a lies behind m and is eventually mapped to an endpoint of m under σ_2 then there is a minor $m_a \ni a$ that is eventually mapped to m under σ_2.

The first claim of Theorem A easily implies the second one.

Renormalization and baby QMLs. The *empty* lamination is the lamination all of whose leaves are degenerate (i.e., are singletons in $\mathbb{S}$).

Consider two quadratic laminations $\mathcal{L}_1$ and $\mathcal{L}_2$. If $\mathcal{L}_2 \subset \mathcal{L}_1$, then we say that $\mathcal{L}_1$ *tunes* $\mathcal{L}_2$; in particular this means that any lamination trivially tunes itself. If $\mathcal{L}_2 \subsetneqq \mathcal{L}_1$, then $\mathcal{L}_1$ is obtained out of $\mathcal{L}_2$ by inserting some chords (which become leaves of $\mathcal{L}_1$) in gaps of $\mathcal{L}_2$. If in this setting $\mathcal{L}_2 = \mathcal{L}(m_2)$ for a non-degenerate periodic minor m_2 (we do not exclude the possibility $\mathcal{L}_2 = \mathcal{L}_1$), then $\mathcal{L}_1$ is called *renormalizable*. A lamination $\mathcal{L}_1$ is *almost non-renormalizable* if there exists no non-empty lamination $\mathcal{L}(m_2) \subsetneqq \mathcal{L}_1$. We call $\mathcal{L}_1$ **almost** non-renormalizable because if it is as above while also $\mathcal{L}_1 = \mathcal{L}(m_1)$ with non-degenerate periodic minor m_1 then, as we saw above, $\mathcal{L}_1$ is renormalizable, but only in a trivial way. Observe that in [**BOT17**] almost non-renormalizable laminations are called *oldest ancestors*.

Let m be a non-degenerate periodic minor. We will write $C(m)$ for the *central set* of $\mathcal{L}(m)$, i.e., the gap/leaf of $\mathcal{L}(m)$ containing the center of $\mathbb{D}$ and, therefore, located between the two majors of $\mathcal{L}(m)$. Equivalently, $C(m)$ can be called the *critical* set of $\mathcal{L}(m)$. Then $\sigma_2(C(m))$ is the convex hull of $\sigma_2(\mathbb{S} \cap C(m))$. This is also a gap or a leaf of $\mathcal{L}(m)$ having m as a boundary leaf (edge). We will see that, if $\mathcal{L}(m_1)$ is renormalizable, then m_1 is contained in $\sigma_2(C(m))$ for some m as above. Moreover, we can choose m so that $\mathcal{L}(m)$ is almost non-renormalizable.

All *edges* (i.e., boundary chords) of $\sigma_2(C(m))$ are leaves of QML. However, there are also leaves of QML in $\sigma_2(C(m))$ that enter the interior of $\sigma_2(C(m))$. All these leaves are precisely the minors of all laminations strictly containing $\mathcal{L}(m)$. It follows that all renormalizable laminations are represented by minors in gaps of the form $\sigma_2(C(m))$, where m is periodic and such that $\mathcal{L}(m)$ is almost non-renormalizable. In other words, all minors of almost non-renormalizable laminations and all points in $\mathbb{S}$ form a lamination QML^{nr} ("nr" from **n**on-**r**enormalizable) whose infinite gaps are a special gap CA^c and gaps of the form $\sigma_2(C(m))$, where m is a minor such that $\mathcal{L}(m)$ is almost non-renormalizable. (There are also finite gaps of QML^{nr}; each such gap is a gap of QML too, associated to a non-renormalizable lamination.) Observe that for any periodic minor m the edges of the set $\sigma_2(C(m))$ are leaves of QML (they are minors of laminations that tune $\mathcal{L}(m)$). The gap CA^c, the *combinatorial main cardioid*, is the central gap of QML^{nr} (and of QML itself). By definition, it is bounded by all periodic minors m, for which $\mathcal{L}(m)$ has an *invariant* finite gap adjacent to m, or m is an *invariant* leaf of $\mathcal{L}(m)$. There are no leaves of QML in CA^c, except for the edges of CA^c. The lamination QML^{nr} was introduced in [**BOT17**].

Consider a gap $\sigma_2(C(m))$ of QML^{nr}, where m is a non-degenerate periodic minor (then $\mathcal{L}(m)$ is almost non-renormalizable). Observe that $\sigma_2(C(m))$ is invariant under σ_2^p, where p is the (minimal) period of m. There is a monotone map ξ_m from the boundary of $\sigma_2(C(m))$ to $\mathbb{S}$ that collapses all edges of $\sigma_2(C(m))$. We may also arrange that ξ_m semi-conjugates σ_2^p restricted to the boundary of $\sigma_2(C(m))$ with σ_2. Under ξ_m, any leaf $ab \in$ QML lying in $\sigma_2(C(m))$ is mapped to a leaf $\xi_m(ab) = \xi_m(a)\xi_m(b)$ of QML. In this sense, we say that leaves of QML lying in $\sigma_2(C(m))$ form a *baby* QML. Thus, QML admits the following *self-similar description*: the lamination QML is the union of QML^{nr} and all baby QMLs inserted in infinite gaps of the form $\sigma_2(C(m))$.

To complete this self-similar description we suggest an explicit construction for QML^{nr} in terms of offsprings.

THEOREM B. The lamination QML^{nr} is obtained as the set of all offsprings of the edges $m \subset \mathrm{CA}^c$ and the limits of such offsprings.

Theorem B parallels the encoding of the Mandelbrot set in terms of "the Yoccoz combinatorial analytic invariants" introduced by C. Petersen and P. Roesch in [**PR08**], more specifically see Corollary 3.23 from [**PR08**] (we are indebted to one of the referees for this remark).

Dynamical generation of the QML**.** Theorem B is the basis for a dynamical generation of the QML. The construction consists of three steps repeated countably many times, and then one final step.

Step 1. First, we construct all edges of the combinatorial main cardioid. The endpoints of these edges can be computed explicitly.

Step 2. For every edge m of the CA^c, we construct all offsprings of m. As follows from Theorem B, taking offsprings is as easy as taking pullbacks of a leaf in an invariant lamination.

Step 3. Take the limits of all offsprings from step 2. We obtain a lamination behind m with gaps of the form $\sigma_2(C(m_1))$, where m_1 is a periodic minor behind m such that $\mathcal{L}(m_1)$ is almost non-renormalizable. Drawing these laminations for all edges of CA^c gives the lamination QML^{nr}.

Step 4. In each gap of the form $\sigma_2(C(m_1))$ as above, construct chords whose ξ_{m_1}-images are leaves constructed at steps 1–3. In other words, we repeat our construction for each baby QML, and then keep repeating it countably many times. Let us denote the thus obtained family of leaves of QML by QML^{fr}. By [**BOT17**], QML^{fr} includes all minors of so-called *finitely renormalizable* quadratic laminations ("fr" comes from "finitely renormalizable") so that the only minors that are missing are the ones that correspond to *infinitely renormalizable* laminations, i.e. laminations $\mathcal{L}$ for which there exists a nested infinite sequence of pairwise distinct laminations $\mathcal{L}_1 \subset \mathcal{L}_2 \subset \dots$ such that $\mathcal{L}_n \subset \mathcal{L}$ for any n.

Step 5. To get the missing minors we now take the limits of leaves of QML^{fr}. Notice that, by [**BOT17**], these limit minors are, for the most part, degenerate (i.e., they are singletons in $\mathbb{S}$). The limit minors that are non-degenerate are exactly those that correspond to the quadratic laminations $\mathcal{L}(m)$ that are infinitely renormalizable with the following additional property: $\mathcal{L}(m)$ coincides with a q-lamination $\mathcal{L}_{\sim_m}$ associated to a laminational equivalence $\sim_m$ such that the corresponding topological polynomial contains a periodic arc in its topological Julia set.

Possible applications to other parameter slices. The problem of constructing models of the entire connectedness locus in degrees greater than 2 seems to be rather complicated. Even in degree three there are no known "global" models of this space. In this brief discussion we will, therefore, talk about complex one dimensional slices of parameter spaces of higher degree polynomials; moreover, for the sake of simplicity we will only deal with the cubic case. Finally, for the sake of brevity we assume familiarity with basic concepts of combinatorial complex dynamics.

One of the main goals of this paper is to develop tools and techniques that can be used to construct combinatorial models for complex one dimensional slices of parameter spaces of cubic polynomials. Indeed, by C. McMullen [**McM00**], slices of the cubic connectedness locus contain lots of copies of $\mathcal{M}_2$ to which our results apply directly (in fact, the article [**McM00**] contains much more general results).

However otherwise the situation is not as simple. A lot of results show that in the cubic case various parameter slices are *not* locally connected. Lavaurs [**Lav89**] proved that the cubic connectedness locus itself is not locally connected. Epstein and Yampolsky [**EY99**] showed that the bifurcation locus in the space of real cubic polynomials is not locally connected either. Buff and Henriksen [**BH01**] presented copies of quadratic Julia sets, including not locally connected Julia sets, in slices of $\mathcal{M}_3$. These are complications of analytic and topological nature.

There are also combinatorial hurdles that need to be overcome. To begin with, Thurston's Central Strip Lemma 1.4 fails already in the cubic case; e.g., if a cubic lamination admits a critical quadrilateral Q associated with the critical strip S, and a critical leaf ℓ, then the forward orbit of Q may come close to ℓ and then enter S, a dynamical phenomenon impossible in the quadratic case because of the Central Strip Lemma. In addition, Thurston's No Wandering Triangle Theorem (Theorem 1.11) also fails in the cubic case [**BO04**, **BO08**]. This complicates *both* the task of constructing a combinatorial model of slices of cubic polynomial spaces *and* the task of applying the idea of the present paper to such slices even assuming that the laminational model for (some) slices have been constructed.

In fact, we are not aware of many combinatorial models of such spaces (even though we believe that a lot of them admit combinatorial models in terms of laminations). An example one might consider is given in the paper [**BOPT16c**] which we now discuss. Consider the tripling map $\sigma_3 : \mathbb{S} \to \mathbb{S}$ and fix a critical leaf D of σ_3. Moreover, choose D so that it cannot be a boundary leaf of a periodic Siegel gap. Then consider the space of all cubic laminational equivalence relations $\sim$ which have a critical class containing the endpoints of D (e.g., the endpoints of D may well be a class of this equivalence relation). Observe that in this case the class containing the endpoints of D must be finite.

To each such equivalence relation $\sim$ we associate its *minor set* $m_\sim$ defined as follows. First, if there is a unique critical set (class) of $\sim$, then $m_\sim$ is the convex hull of its image. Second, if there are two $\sim$-classes and both are finite, then we choose the one not containing the endpoints of D and set $m_\sim$ to be the convex hull of the image of this $\sim$-class. Finally, consider the remaining case which is as follows: $\sim$ has a unique periodic *critical* Fatou gap of period k such that $\sigma_3^k : U \to \sigma_3(U)$ is two-to-one. Evidently, this implies that $\sigma_3^k : U \to U$ is two-to-one. We show in [**BOPT16c**] that there is a unique edge $M_\sim$ of U of period k. In this case we set $m_\sim = \sigma_3(M_\sim)$.

One of the main results of [**BOPT16c**] is that the minor sets $m_\sim$ can be viewed as tags of their laminational equivalence relations $\sim_D$ (so that the space of all such laminational equivalence relations is similar to $\mathcal{M}_2$) while the collection of their convex hulls will give rise to a lamination $\mathcal{L}_D$. The corresponding space of all cubic laminations that admit critical leaf D is $\mathbb{S}/\sim_D$. We hope that the ideas and results of this paper can be properly adjusted to lead to a more explicit description of the structure of $\mathbb{S}/\sim_D$ at least for some critical leaves D. A likely candidate for that is the critical leaf $D = \overline{\frac{1}{3}\frac{2}{3}}$, first preimage of a σ_3-fixed angle 0. This is based upon the fact that if $D = \overline{\frac{1}{3}\frac{2}{3}}$, then we can prove the Central Strip Lemma for all laminations admitting D, and this allows us to apply similar arguments to the present paper, in particular concerning pulling back the minors and thus constructing new minors.

In general, the plan can be as follows. Consider a parameter slice and assume that its combinatorial model exists. This model will be a lamination $\mathcal{L}$ in $\mathbb{D}$. In order to construct $\mathcal{L}$, we will apply a similar procedure to the one described above for QML. Steps 1–3 will be replaced with similar steps. However, step 4 will operate with genuine baby QMLs rather than copies of $\mathcal{L}$. Thus, the lamination $\mathcal{L}$ will consist of a sublamination $\mathcal{L}^{nr}$ in whose infinite gaps we insert copies of QML rather then copies of $\mathcal{L}^{nr}$ itself.

Evidently, a lot of details in the actual implementation of the outlined approach will be very different from what is done in the current paper. There are also complications related to the fact that some quadratic techniques fail for higher-degree polynomials. Instead of Thurston's technique based on the Central Strip Lemma, we will have to rely on methods developed in [**BOPT16c**] or, more generally, in [**BOPT17**]. However, even in the simplest cases of cubic parameter laminations, a complete implementation of this program will require at least as much space as this paper. Thus we postpone the details to future publications. Still, we believe that the sketched technique should (hopefully!) work for some (but not all) complex one dimensional slices.

To summarize, we think that while our dynamical approach to the construction of the Mandelbrot set is quite consistent with the more static viewpoints of Thurston [**Thu85**], Keller [**Kel00**], Lavaurs [**Lav86**, **Lav89**], and Schleicher [**Sch09**], it is based upon a familiar pullback construction which has its own advantages, in particular making it more accessible to those familiar with that dynamically-based process.

1. Majors and minors

In this section, we recall fundamental properties of quadratic laminations. Since all statements here can be traced back to [**Thu85**], we skip references to this seminal paper of Thurston until the end of the section (see also [**Sch09**] and [**BOPT16**] where some of these results are more fleshed out). The exposition is adapted to our purposes, and some facts are stated in a different but equivalent form (see [**BMOV13**] for an extension of this approach to higher degree laminations). Some proofs are omitted.

1.1. Notation and terminology. As usual, $\mathbb{C}$ is the plane of complex numbers identified with the real 2-dimensional vector space $\mathbb{R}^2$. For any subset $A \subset \mathbb{C}$, we let $\overline{A}$ denote its closure. For any set $G \subset \overline{\mathbb{D}}$ of the form $G = \mathrm{CH}(G \cap \mathbb{S})$, we let $\sigma_2(G)$ denote the set $\mathrm{CH}(\sigma_2(G \cap \mathbb{S}))$. Chords of $\mathbb{S}$ on the boundary of G are called *edges* of G. A chord of $\mathbb{S}$ with endpoints a, $b \in \mathbb{S}$ is denoted by ab. If $a = b$, then the chord is said to be *degenerate*, otherwise it is said to be *non-degenerate*.

We will identify $\mathbb{R}/\mathbb{Z}$ with $\mathbb{S}$ by means of the map $\theta \in \mathbb{R}/\mathbb{Z} \mapsto \overline{\theta} = e^{2\pi i\theta}$. Elements of $\mathbb{R}/\mathbb{Z}$ are called *angles*. The point $\overline{\theta}$ will be sometimes referred to as the point in $\mathbb{S}$ of angle θ. For example $\overline{0}$ and $\overline{\frac{1}{2}}$ are the only points of $\mathbb{S}$ lying on the real axis, and $\overline{0\frac{1}{2}}$ is the corresponding diameter. In order to avoid confusion, we will always write $\overline{0}$, $\overline{\frac{1}{2}}$, $\overline{\frac{1}{4}}$ rather than 1, -1, i, etc.

Let M be a chord of the unit circle. We will write $-M$ for the chord obtained from M by a half-turn, i.e., by the involution $z \mapsto -z$. Let S be the (closed) strip between M and $-M$. Define the map $\psi : [0, \frac{1}{2}] \to [0, \frac{1}{2}]$ by $\psi(x) = 2x$ if $0 \leqslant x \leqslant \frac{1}{4}$ and $\psi(x) = 1 - 2x$ if $\frac{1}{4} \leqslant x \leqslant \frac{1}{2}$; the fixed points of ψ are 0 and $\frac{1}{3}$. Then it is easy to see that given a chord ℓ, we have $|\sigma_2(\ell)| = \psi(|\ell|)$. The dynamics of ψ shows that for any non-degenerate chord ℓ there exists $n \geqslant 0$ such that $|\sigma_2^n(\ell)| \geqslant \frac{1}{3}$. Hence if $\pm M$ are the majors of a lamination then $|M| \geqslant \frac{1}{3}$ and $|\sigma_2(M)| \leqslant \frac{1}{3}$. Suppose that $|M| \geqslant \frac{1}{3}$ and that $m = \sigma_2(M)$ is disjoint from the interior of S. Then the chords $\pm M$ and the strip S are uniquely determined by m. Under the assumptions just made, we call m *minor-like*, set $S = S(m)$, and call it the *central strip* of m. Observe that if m is degenerate, then $S(m) = M = -M$ is a diameter, in particular, it has no interior. We will write $Q(m)$ for the quadrilateral $\mathrm{CH}(M \cup (-M))$.

LEMMA 1.1. *Suppose that $\ell = ab, a \neq b$ is a leaf of a lamination $\mathcal{L}$ such that $|\ell| \leqslant \frac{1}{3}$ and $\overline{0} \notin \overline{H(\ell)}$. Then ℓ is minor-like. In particular if m is a minor and $\ell \leqslant m$, then ℓ is minor-like.*

PROOF. Either two or all four edges of $Q(\ell)$ are leaves of $\mathcal{L}$. If only one vertex of $Q(\ell)$ belongs to $H(\ell)$, then at least one edge of $Q(\ell)$ belongs to $\mathcal{L}$ and crosses ℓ, a contradiction. Hence either two preimages a', b' of points a and b, respectively, belong to $H(\ell)$, or none. Set $\ell' = a'b'$; then $\sigma_2(\ell') = \ell$. Suppose that $\ell' \neq \ell$; then $|\ell'| < |\ell|$. If $\sigma_2(H(\ell'))$ is $\mathbb{S} \setminus \overline{H(\ell)}$ then the fact that $|\ell| \leqslant \frac{1}{3}$ implies that $|\mathbb{S} \setminus \overline{H(\ell)}| \geqslant \frac{2}{3}$ and hence $|H(\ell')| \geqslant \frac{1}{3}$, a contradiction with $|\ell'| < |\ell| \leqslant \frac{1}{3}$. Hence $\sigma_2(H(\ell')) = H(\ell)$ and $|\ell'| \leqslant \frac{1}{6}$. Moreover, the restriction of σ_2 to $H(\ell')$ is one-to-one and expanding. It follows that σ_2 has a fixed point in $H(\ell')$. The only fixed point of σ_2 is $\overline{0}$, hence we have $\overline{0} \in \overline{H(\ell')}$, a contradiction. Thus, either $\ell' = \ell$ or $Q(\ell) \cap H(\ell) = \varnothing$ (evidently, all vertices of $Q(\ell)$ cannot belong to $\overline{H(\ell)}$). In the former case it follows that $\ell = \overline{\frac{1}{3}\frac{2}{3}}$ is a minor, in the latter case ℓ is minor-like by definition. For the last claim of the lemma, note that if m is a minor, then $\overline{0} \notin \overline{H(m)}$. □

A *critical chord* is a diameter of $\mathbb{S}$. The endpoints of a critical chord are mapped under σ_2 to the same point of $\mathbb{S}$. A set $G \subset \overline{\mathbb{D}}$ of the form $G = \mathrm{CH}(G \cap \mathbb{S})$ is said to be *semi-critical* if G contains a critical chord. Equivalently, a semi-critical set contains the center of the disk.

1.2. The Central Strip Lemma. A chord of $\mathbb{S}$ is said to be *vertical* if it separates $\overline{0}$ from $\overline{\frac{1}{2}}$, and *horizontal* otherwise. The distinction between the two types of chords is important for quadratic laminations.

LEMMA 1.2. *Let m be a minor-like chord. Then $\sigma_2(S(m) \cap \mathbb{S}) = \overline{H(m)}$.*

PROOF. The set $S(m) \cap \mathbb{S}$ consists of two arcs, each of length $\leqslant \frac{1}{6}$. Both arcs map to the same arc A of length $\leqslant \frac{1}{3} < \frac{1}{2}$. On the other hand, A is bounded by the endpoints of m, hence $A = \overline{H(m)}$. □

LEMMA 1.3. *Let m be a non-degenerate minor-like chord. Then $S(m)$ is bounded by vertical chords.*

The only degenerate minor-like chord for which the statement fails, is $\overline{0}$.

PROOF. Assume that the edges $\pm M$ of $S(m)$ are horizontal. Then $\overline{0\frac{1}{2}} \subset S(m)$, hence, $\overline{0} \in \overline{H(m)}$ by Lemma 1.2. Thus $\overline{0}$ belongs to both $S(m)$ and $\overline{D(m)}$. On the other hand, by definition of minor-like chords, these two sets cannot have common interior points. It follows that $\overline{0}$ is an endpoint of m. Let $\overline{\alpha}$ be the other endpoint. Then $\overline{\frac{\alpha}{2}} \in H(m) \cap S(m)$, a contradiction. □

Let us make the following observations.

(a) If ℓ is a chord of $\mathbb{S}$ such that $|\ell| \leqslant \frac{1}{4}$, then $|\sigma_2(\ell)| = 2|\ell|$; otherwise $|\sigma_2(\ell)| = 1 - 2|\ell|$.
(b) We have $|\sigma_2(\ell)| > |\ell|$ if and only if $|\ell| < \frac{1}{3}$.
(c) If ℓ is disjoint from the edges $\pm M$ of $S(m)$ and $|\ell| > |M|$, then ℓ is a vertical chord in $S(m)$ (here $m = \sigma_2(M)$ is minor-like).
(d) Any non-degenerate chord eventually maps to a chord of length $\geqslant \frac{1}{3}$.

Lemma 1.4 (The Central Strip Lemma). *Let m be a minor-like chord. Suppose that the chords $\sigma_2^n(m)$ do not cross any edge of $S(m)$ for any $n > 0$.*

(1) *If $|\sigma_2^n(m)| < |m|$ for a minimal $n > 0$, then $\sigma_2^{n-1}(m)$ is a vertical chord in $S(m)$ distinct from either edge of $S(m)$, and $\sigma_2^n(m) \leqslant m$;*
(2) *if $\sigma_2^n(m) \subset S(m)$ for some $n > 0$, and n is the smallest positive integer with this property, then the chord $\sigma_2^n(m)$ is vertical.*

Proof. We will write $\pm M$ for the edges of $S(m)$. To prove (1), observe that $|\sigma_2^{n-1}(m)| > |M|$ which implies that $\sigma_2^{n-1}(m) \subset S(m)$ is a vertical chord. Observe now that (2) follows from (1) since if $\sigma_2^n(m) \subset S(m)$ is horizontal, then $|\sigma_2^n(m)| \leqslant |m|/2$. □

1.3. Minor leaf laminations. By definition, the Central Strip Lemma, and by observations (a) – (d), a minor m has the following properties:

(SA1) it is minor-like;
(SA2) all $\sigma_2^n(m)$, where $n \geqslant 0$, are pairwise unlinked and do not cross any edges of $S(m)$;
(SA3) for any $n > 0$ we have $|\sigma_2^n(m)| \geqslant |m|$;
(SA4) if $\sigma_2^n(m) \leqslant m$ for some $n > 0$, then $\sigma_2^n(m) = m$ (thus, images of m are disjoint from $D(m) \cup S(m) \setminus (M \cup -M)$).

For brevity, in what follows we will refer to these properties simply as SA1, SA2, SA3 and SA4. Clearly, SA3 always implies SA4. Moreover, by the Central Strip Lemma, if SA1 and SA2 hold for a chord m, then SA3 and SA4 for this chord are equivalent.

Definition 1.5 (Stand Alone Minor). A chord m is called a *stand alone minor* if properties SA1–SA3 hold. (Then automatically SA4 also holds).

Note that all points of $\mathbb{S}$ are stand alone minors. Any stand alone minor is the minor of a certain quadratic lamination. Any such lamination can be constructed by "pulling back" the minor and all its images. Such pullbacks are mostly unique but, if m is periodic, allow for small variations.

In this paper we establish *dynamical conditions* that imply that certain leaves of a lamination $\mathcal{L}$ with minor m are minors themselves. We do this by verifying for them that they are stand alone minors. This requires checking for them conditions SA1 – SA3. It turns out that depending on the location of ℓ with respect to m or the length of ℓ with respect to the length of m some of these conditions easily follow.

LEMMA 1.6. *Let ℓ be a leaf of a lamination $\mathcal{L}$ with minor m. Then the following holds.*

(1) *Choose the least $i \geqslant 0$ with $|\sigma_2^i(\ell)| \geqslant |m|$. Then $|\sigma_2^j(\ell)| \geqslant |m|$ for any $j \geqslant i$. Thus, if $|\ell| \leqslant |m|$ then $|\ell| \leqslant |\sigma_2(\ell)| \leqslant \dots \leqslant |\sigma_2^i(\ell)|$ so that property* SA3 *holds for ℓ. In particular, $\sigma_2^t(m) \leqslant m, \sigma_2^t(m) \neq m$ is impossible.*

(2) *If $m \leqslant \ell$, then no eventual image of ℓ crosses the edges of $S(\ell)$ so that property* SA2 *holds for ℓ.*

PROOF. (1) By assumption, $|\sigma_2^i(\ell)| \geqslant |m|$. If $|\sigma_2^j(\ell)| < |m|$ for some $j \geqslant i$, then, by the Central Strip Lemma, for some k the leaf $\sigma_2^k(\ell)$ is vertical inside $\overline{S(m)}$, a contradiction with the vertical pullbacks $\pm M$ of m being the majors of $\mathcal{L}$. Observe that $|m| \leqslant \frac{1}{3}$ as was explained in the paragraph right before Lemma 1.1. Hence for each $r, 0 \leqslant r \leqslant i-1$ we have $|\sigma_2^r(\ell)| \leqslant |m| \leqslant \frac{1}{3}$ which easily implies that $|\sigma_2^r(\ell)| \leqslant |\sigma_2^{r+1}(\ell)|$, $r = 0, \dots, i-1$.

(2) Since the horizontal pullbacks of ℓ cross the vertical edges of $S(m)$, which are leaves of $\mathcal{L}$, the vertical pullbacks $\pm L$ of ℓ (which are the edges of $S(\ell)$) must be leaves of $\mathcal{L}$. Hence eventual images of ℓ do not cross an edge of $S(\ell)$, as desired. □

A few well-known results concerning quadratic laminations with a given minor m are summarized in Theorem 1.7; these results can be found in [**Thu85**], or can be easily deduced from [**Thu85**].

THEOREM 1.7. *If m is a stand alone minor, then there exists a quadratic lamination $\mathcal{L}$ with minor m. Depending on m, the following holds.*

(1) *If m is non-periodic, then either*
   (a) *a quadratic lamination $\mathcal{L}$ with minor m is unique, or*
   (b) *if in addition m is non-degenerate, then there are at most two quadratic laminations $\widehat{\mathcal{L}} \subset \mathcal{L}$ with minor m one of which must be a q-lamination $\widehat{\mathcal{L}}$ with finite gaps.*

(2) *If m is periodic and non-degenerate, then there exists a unique q-lamination $\mathcal{L}$ such that m is its minor.*

(3) *If m is periodic and degenerate, then there are at most four quadratic laminations with m as a minor, and there exists a unique q-lamination $\widehat{\mathcal{L}}$ whose periodic minor $\hat{m}$ has m as an endpoint. Moreover, if $m \neq \overline{0}$ then $\hat{m}$ is non-degenerate.*

In any case, there exists a unique q-lamination $\widehat{\mathcal{L}}(m)$ such that, if m is not a periodic point, then any lamination with minor m contains $\widehat{\mathcal{L}}(m)$; moreover, if m is non-degenerate and non-periodic, then all leaves of $\widehat{\mathcal{L}}(m)$ are non-isolated in $\widehat{\mathcal{L}}(m)$ and all gaps of $\widehat{\mathcal{L}}(m)$ are finite. In case (1)(b)*, any leaf of $\mathcal{L} \setminus \widehat{\mathcal{L}}(m)$ is eventually mapped to vertical edges of $Q(m)$.*

We can now define a specific lamination $\mathcal{L}(m)$ with minor m.

DEFINITION 1.8. If m is a non-periodic or non-degenerate stand alone minor, define $\mathcal{L}(m)$ as one of the laminations from Theorem 1.7 as follows: in case (1)(a) the lamination $\mathcal{L}(m)$ is the unique quadratic lamination with minor m; in case (1)(b), the lamination $\mathcal{L}(m)$ is the bigger of the two laminations with minor m; in case (2) it is the unique q-lamination with minor m. In any case the central set of $\mathcal{L}(m)$ is denoted by $C(m)$. Finally, the q-lamination $\widehat{\mathcal{L}}$ from Theorem 1.7 will be denoted by $\widehat{\mathcal{L}}(m)$ and will be called the *q-lamination associated with m.*

This defines $\mathcal{L}(m)$ except for the case when m is a periodic singleton (which will be done later). By definition, m is the minor of $\mathcal{L}(m)$. Observe that the central set $C(m)$ of a lamination $\mathcal{L}(m)$ is either a critical leaf (a diameter), a collapsing quadrilateral, or an infinite periodic quadratic gap.

In the sequel, by a *minor* we mean a stand alone minor or, which is the same by Theorem 1.7, the minor of some (not specified) quadratic lamination. Minors are also identical to leaves of the QML. The lamination $\mathcal{L}(m)$ is called the *minor leaf lamination* associated with a minor m. In order to construct $\mathcal{L}(m)$, we will describe the process of taking pullbacks of chords.

DEFINITION 1.9 (m-pullbacks). Let m be a minor-like chord, let $\ell = ab$ be a chord of $\mathbb{S}$ that is not linked with m. The *m-pullbacks* of ℓ are defined as follows. If $\ell = m$, then the m-pullbacks are the major(s) $\pm M$, the edges of $S(m)$. If $\ell \neq m$ is a point in $\mathbb{S}$, then the m-pullbacks of ℓ are points in $\sigma_2^{-1}(\ell)$. Otherwise, there are four points in $\sigma_2^{-1}(\ell \cap \mathbb{S})$, and there are two possible cases. First, $\ell \subset \overline{D(m)}$ in which case all four points belong to $S(m)$. Then we define the m-pullbacks of ℓ as the horizontal pullbacks of ℓ. Second, $\ell \subset \overline{\mathbb{D}} \setminus D(m)$ in which case all four points belong to $\overline{\mathbb{S} \setminus S(m)}$. If m is non-degenerate or ℓ is disjoint from m, the m-pullbacks of ℓ are defined as the two pullbacks of ℓ that do not cross M or $-M$. In the remaining case m is degenerate and is an endpoint of ℓ; then we define the m-pullbacks of ℓ to be the pullbacks of ℓ that have length $\leqslant \frac{1}{4}$.

In the last case in Definition 1.9, if $m \neq \overline{0}$ or if $m = \overline{0}$ but $\ell \neq \overline{0\frac{1}{2}}$, there are exactly two m-pullbacks of ℓ while if $m = \overline{0}$ (hence $M = \overline{0\frac{1}{2}}$) and $\ell = M$ then there are four such pullbacks: $\overline{0\frac{1}{4}}$, $\overline{\frac{1}{4}\frac{1}{2}}$, $\overline{\frac{1}{2}\frac{3}{4}}$ and $\overline{\frac{3}{4}0}$.

Observe that if (degenerate) $m \neq \overline{0}$ is an endpoint of ℓ then the m-pullbacks of ℓ are horizontal. Indeed, in that case $M \neq \overline{0\frac{1}{2}}$ is a diameter of $\mathbb{D}$ with endpoints $\pm a$. We may assume that a is in the upper half-plane. Then $m = \sigma_2(a) < M$. If ℓ is small, then the m-pullbacks of ℓ are two short chords coming out of the points $\pm a$ (the other two candidate pullbacks are of length $> \frac{1}{4}$). Clearly, both chords are horizontal. As we continuously increase the length of ℓ, its pullbacks also continuously increase. The longest option for ℓ is still shorter than a half-circle, hence these chords are m-pullbacks of ℓ. If at some moment they stop being horizontal, then at this moment the endpoints of these chords not in M must become either $\overline{0}$ or $\overline{\frac{1}{2}}$. Hence their common image ℓ must have an endpoint $\sigma_2(\overline{0}) = \overline{0}$. However ℓ cannot have $\overline{0}$ as an endpoint, a contradiction.

Importantly, there is no way of making m-pullbacks depend continuously on m. This is why the definition of m-pullbacks may not look very natural. Observe the following. If m is a minor, then any chord of the form $\sigma_2^n(m)$ is an m-pullback of $\sigma_2^{n+1}(m)$ for $n \geqslant 0$. Indeed, this statement is non-trivial only for non-degenerate m. In this case m-pullbacks are determined by the property that they do not cross the edges of $S(m)$ (by property (4) of minors, iterated images of m never enter $S(m) \setminus (M \cup -M)$). The following theorem complements Theorem 1.7; recall that in case when m is non-degenerate, or degenerate and non-periodic, $\mathcal{L}(m)$ was defined above (see Theorem 1.7).

THEOREM 1.10. *If m is a non-degenerate or non-periodic minor then iterated m-pullbacks of iterated σ_2-images of m are dense in $\mathcal{L}(m)$.*

In fact, Theorem 1.10 inspires the *definition* of $\mathcal{L}(m)$ in the only remaining case when m is a periodic singleton; in that case we define $\mathcal{L}(m)$ as the closure of the family of all iterated m-pullbacks of M where M is the diameter mapped to m by σ_2.

1.4. Classification of dynamic gaps. The key tool that allowed Thurston to succeed in establishing a complete classification of gaps of quadratic laminations was Theorem 1.11 (No Wandering Triangles Theorem). Let Δ be a triangle with vertices in $\mathbb{S}$. It is said to be *wandering* if all $\sigma_2^n(\Delta)$ have non-empty disjoint interiors for $n \geqslant 0$.

THEOREM 1.11. *Wandering triangles do not exist.*

The first step in the classification of all gaps is the following corollary.

COROLLARY 1.12. *Let G be a gap of a quadratic lamination $\mathcal{L}$. Then an eventual image of G either contains a diameter or is periodic and finite.*

Semi-critical gaps are classified as follows:

- strictly preperiodic critical finite gaps with more than 4 edges;
- *collapsing quadrilaterals*, i.e., quadrilaterals that are mapped to non-degenerate leaves;
- *collapsing triangles*, i.e., triangles with a critical edge;
- *caterpillar gaps*, i.e., periodic gaps with a critical edge.
- *Siegel gaps*, i.e., infinite periodic gaps G such that $G \cap \mathbb{S}$ is a Cantor set, σ_2^n maps G onto itself for some n, and σ_2^n restricted to the boundary of G is semi-conjugate to an irrational rotation of the circle under the map that collapses all edges of G to points.

All edges of a caterpillar gap are eventually mapped to the critical edge. Any caterpillar gap has countably many edges and countably many vertices.

Let $A \subset \mathbb{S}$ be a compact set. Denote by $\sigma_d : \mathbb{S} \to \mathbb{S}$ the d-tupling map that takes z to z^d for any $d \geqslant 2$. We say that $\sigma_d : A \to \sigma_d(A)$ has *degree k covering property* if there is a degree k orientation preserving covering $f : \mathbb{S} \to \mathbb{S}$ such that $\sigma_d|_A = f|_A$ and such k is minimal.

PROPOSITION 1.13. *Consider a gap G of a quadratic lamination $\mathcal{L}$ such that no edge of G is a critical leaf. Then the map $\sigma_2 : G \cap \mathbb{S} \to \sigma_2(G \cap \mathbb{S})$ has degree k covering property, where $k = 1$ or 2.*

A bijection from a finite subset A of $\mathbb{S}$ to itself is a *combinatorial rotation* if it preserves the cyclic order of points. Thus, a combinatorial rotation $f : A \to A$ is a map which extends to an orientation preserving homeomorphism $g : \mathbb{S} \to \mathbb{S}$, topologically conjugate to a Euclidean rotation. A gap G of a quadratic lamination $\mathcal{L}$ is *periodic* if $\sigma_2^p(G) = G$ for some $p > 0$; the smallest such p is the *period* of G. If G is of period p, then σ_2^p restricted to $G \cap \mathbb{S}$ is the *first return map* of G. By Proposition 1.13 the first return map of a finite periodic gap is a combinatorial rotation. Moreover, if $\mathcal{L}$ has no critical leaves then the first return map of an infinite periodic gap G has the degree 2 covering property and $G \cap \mathbb{S}$ is a Cantor set.

LEMMA 1.14. *Let G be a periodic gap of a quadratic lamination $\mathcal{L}$, and $f : G \cap \mathbb{S} \to G \cap \mathbb{S}$ its first return map.*

(1) *If G is finite, then f is a transitive combinatorial rotation. In particular, for any a, $b \in G \cap \mathbb{S}$ such that ab is not an edge of G, the chord $f^k(ab)$ crosses ab for some $k > 0$.*
(2) *If $a \neq b \in \mathbb{S} \cap G$ and neither a nor b eventually maps to $\overline{0}$, then $f^k(ab)$ is vertical for some $k \geqslant 0$. This is true, e.g., if the interior of G contains the center of $\mathbb{D}$, the lamination $\mathcal{L}$ is non-empty, and a, b are arbitrary points in $G \cap \mathbb{S}$.*

PROOF. The only claim that is not explicitly contained in [**Thu85**] is the last one. Assume, by way of contradiction, that $f^k(ab)$ is horizontal for all $k \geqslant 0$. Then, for every k, either both $f^k(a)$ and $f^k(b)$ are in the open upper half of $\mathbb{D}$, or both in the open lower half of $\mathbb{D}$. Suppose that a point $x \in \mathbb{S}$ never maps to $\overline{0}$. Define the *address* of x as U if x is above $\overline{0\frac{1}{2}}$ and as L otherwise. (The symbols U and L come from "Upper" and "Lower"). The *itinerary* of x is an infinite word in the alphabet $\{U, L\}$ consisting of addresses of all $f^k(x)$ for $k \geqslant 0$. Similarly, we can define *finite itineraries* of length N if, instead of all $k \geqslant 0$, we take all k such that $0 \leqslant k < N$. It is easy to see that the locus of points with a given finite itinerary is an arc in $\mathbb{S}$. Moreover, this arc has length 2^{-N}, where N is the length of the itinerary. It follows that every infinite itinerary defines at most one point. In particular, since by the assumption a and b have the same itinerary, we conclude that $a = b$, a contradiction.

If G contains the center of $\mathbb{D}$ in its interior, then $\mathcal{L}$ does not have critical leaves. Hence σ_2 has a degree k covering property on G, with $k = 1$ or 2. We claim that then $\overline{0} \notin G$. Indeed, suppose otherwise. Then it is easy to see that G is invariant and, hence, $f = \sigma_2$. Now, if G is finite, then f fixes $\overline{0}$, hence cannot act as a transitive combinatorial rotation. If G is infinite, then the fact that $\mathcal{L}$ has no critical leaves implies that f has degree 2 covering property on G. Using the density of $\bigcup_{n \geqslant 0} \sigma_2^{-n}(\overline{0})$ in both $\mathbb{S}$ and $G \cap \mathbb{S}$, we conclude that $G = \overline{\mathbb{D}}$ and $\mathcal{L}$ is the empty lamination, a contradiction. Hence we may assume that $\overline{0}$ (and, therefore $\overline{\frac{1}{2}}$) do not belong to G. Since G is periodic, the points $\overline{0}$ and $\overline{\frac{1}{2}}$ do not belong to iterated σ_2-images of G either. This implies that if $a \neq b \in G \cap \mathbb{S}$ then, by the first paragraph, $f^k(ab)$ is vertical for some $k \geqslant 0$. □

1.5. Classification of parameter gaps. Thurston classified all gaps of QML (see Theorem II.6.11 of [**Thu85**]); we outline this classification below.

Suppose first that G is a finite gap of QML. Then G is strictly preperiodic under σ_2. Moreover, it is the σ_2-image of a finite central set C in a quadratic lamination $\mathcal{L}$. The gap C has 6 edges or more. Conversely, if a quadratic lamination $\mathcal{L}$ has a finite central gap C with 6 or more edges, then $\sigma_2(C)$ is a finite gap of QML. To summarize, finite gaps of QML are precisely finite gaps of quadratic q-laminations that are the images of their central gaps.

Suppose now that G is an infinite gap of QML. Then all edges of G are periodic minors. It may be that $G = \mathrm{CA}^c$. Otherwise, there is a unique edge $m_G = m$ of G such that all $\ell \leqslant m$ for any other edge ℓ of G. Then $G \subset \sigma_2(C(m))$. However, only the edge m is on the boundary of $\sigma_2(C(m))$. Other edges of G enter the interior of $\sigma_2(C(m))$.

It is useful to think about G as a copy of CA^c inserted into $\sigma_2(C(m))$. To make this more precise, observe that there is a monotone continuous map $\xi_m : \mathbb{S} \to \mathbb{S}$ with the following properties. Every complementary component of $\sigma_2(C(m))$ in $\mathbb{S}$, together with endpoints of the edge of $\sigma_2(C(m))$ that bounds it, is mapped to one point. The map ξ_m semi-conjugates the restriction $\sigma_2^p|_{\sigma_2(C(m))\cap\mathbb{S}}$ with σ_2. Here p is the period of $\sigma_2(C(m))$. The map ξ_m is almost one-to-one on $\sigma_2(C(m)) \cap \mathbb{S}$ except that it identifies the endpoints of every edge of $\sigma_2(C(m))$. There is a unique map ξ_m with the properties just listed. Then G is a copy of CA^c in the sense that the ξ_m-images of the edges of G are precisely the edges of CA^c. Moreover, ξ_m-pullbacks are well defined for all edges of CA^c. Indeed, no endpoint of an edge of $\sigma_2(C(m))$ has period > 1 under the first return map to $\sigma_2(C(m))$. Note that, as a consequence, the period of m is the smallest among the periods of all edges of G. Other periods are integer multiples of the period of m.

The case of CA^c is somewhat special as this gap is not associated with any minor. Thurston suggested to think of CA^c as being associated with the degenerate minor $\overline{0}$. Indeed, with these understanding, most properties of infinite gaps of QML extend to the case of CA^c.

2. Derived minors, children, and offsprings: proof of Theorem A

Let us begin with a technical lemma.

LEMMA 2.1. *Let ℓ be a leaf of a quadratic lamination $\mathcal{L}$ where either $\mathcal{L} = \mathcal{L}(m)$, and m is not a periodic point, or $\mathcal{L}$ is a q-lamination. Moreover, let $\sigma_2^i(\ell) \cap \sigma_2^j(\ell) \neq \varnothing$ for some $0 \leqslant i < j$. Then $\sigma_2^i(\ell)$ is a periodic leaf. In particular, if $\ell < m$ and $\sigma_2^n(\ell) = m$ for some n, then all leaves $\sigma_2^i(\ell)$ with $\sigma_2^i(\ell) \leqslant m, \sigma_2^i(\ell) \neq m$, are pairwise disjoint.*

PROOF. By Definition 1.8, the lamination $\mathcal{L}(m)$ is either a q-lamination, or a tuning of a q-lamination with finite gaps. Thus, either $\sigma_2^i(\ell) = \sigma_2^j(\ell)$ is a periodic leaf (mapped to itself under $\sigma_2^{|j-i|}$), or both leaves $\sigma_2^i(\ell)$ and $\sigma_2^j(\ell)$ are contained in the same finite periodic gap G of some q-lamination. However, in the latter case, neither leaf in question can be a diagonal of G because, by Lemma 1.14, eventual images of such diagonals cross each other. Thus again $\sigma_2^i(\ell)$ is a periodic leaf.

Now, let $\ell < m$, set n to be the smallest number such that $\sigma_2^n(\ell) = m$, and assume that $\sigma_2^i(\ell) \cap \sigma_2^j(\ell) \neq \varnothing$ for some $0 \leqslant i < j \leqslant n$. Then, by the above, $\sigma_2^i(\ell)$ and m belong to the same periodic orbit of leaves. However, $\sigma_2^r(m) \leqslant m$ is impossible unless $\sigma_2^r(m) = m$, by Lemma 1.6 . □

Let us now describe several ways of producing new minors ℓ from old ones, cf. part (a) of Lemma II.6.10a in [**Thu85**]. We say that a leaf ℓ *separates* the leaf ℓ' from the leaf ℓ'' if ℓ' and ℓ'' are contained in distinct components of $\mathbb{D} \setminus \ell$ (except, possibly, for endpoints). In particular, this means that $\ell \neq \ell'$ and $\ell \neq \ell''$.

DEFINITION 2.2 (Derived minors and children). Let m be a minor. Let $m_1 \leqslant m$ be a leaf of $\mathcal{L}(m)$ such that eventual images of m_1 do not separate m_1 from m and never equal a horizontal edge of the critical quadrilateral $Q(m)$. Then m_1 is called a *(from m) derived minor*. If, in addition, m_1 is mapped onto m under a suitable iterate of σ_2, then m_1 is called a *child* of m.

By Proposition 2.3 proved below, every derived minor is a minor, justifying its name. If the central gap $C(m)$ of $\mathcal{L}(m)$ is distinct from $Q(m) = \mathrm{CH}(M \cup (-M))$

where M is a major of $\mathcal{L}(m)$ (i.e., if the horizontal edges of $Q(m)$ are not leaves of $\mathcal{L}(m)$), then automatically no image of m_1 equals a horizontal edge of the critical quadrilateral $Q(m)$. Observe, that if $\ell \leqslant m$ and n is the minimal number such that $\sigma_2^n(\ell) = m$, then to verify that ℓ is a from m derived minor it suffices to verify that ℓ never maps to a horizontal edge of $Q(m)$ and that $\sigma_2^i(\ell)$ does not separate ℓ from m for $0 < i < n$ (for $i \geqslant n$ this will hold automatically by Lemma 1.6).

PROPOSITION 2.3. *Let m be a non-degenerate minor. If a leaf $m_1 \in \mathcal{L}(m)$ is a from m derived minor, then m_1 is a minor. Moreover, the horizontal edges of the collapsing quadrilateral $Q(m_1)$ belong to $\mathcal{L}(m)$, and if $\sigma_2^n(m_1) = m$ is the first time m_1 maps to m, then $\sigma_2^{n-1}(m_1)$ is a major of $\mathcal{L}(m)$.*

PROOF. By Lemma 1.1, the chord m_1 is minor-like, i.e., SA1 holds. Let us now check SA2. Since m_1 is a leaf of $\mathcal{L}$, the chords $\sigma_2^k(m_1)$ are unlinked for $k \geqslant 0$. By way of contradiction, suppose that for some $k \geqslant 0$ the chord $\sigma_2^k(m_1)$ crosses an edge M_1 of $S(m_1)$. Then it crosses the edge $-M_1$ since otherwise $\sigma_2^{k+1}(m_1)$ would cross m_1. On the other hand, we know that $\sigma_2^k(m_1)$ cannot cross edges of $S(m)$, hence $\sigma_2^k(m_1) \subset S(m)$. Since $\sigma_2^k(m_1)$ is a leaf of $\mathcal{L}(m)$, it cannot be vertical. Thus $\sigma_2^k(m_1)$ is horizontal and separates the two horizontal edges of $Q(m_1)$. However, this implies that $\sigma_2^{k+1}(m_1)$ separates m_1 from m. A contradiction with the assumption that m_1 is a derived minor.

Property SA3 follows from Lemma 1.6. To prove the next to the last claim, observe that m_1 must have two pullbacks in $\mathcal{L}(m)$, and its vertical pullbacks cannot be leaves of $\mathcal{L}(m)$ as they are longer than the majors of $\mathcal{L}(m)$. The last claim follows from the definition of a derived minor. □

Next we prove a simple but useful technical lemma.

LEMMA 2.4. *The following facts hold.*

1. *If m is a minor, ℓ is a chord such that $\sigma_2^k(\ell) = m$ with k minimal, and $\sigma_2^i(\ell)$ is a horizontal edge of $Q(m)$, then $i = k - 1$.*
2. *If m' and m'' are two distinct non-disjoint minors, then they are edges of the same finite gap G of* QML. *The gap G is the image of a finite critical gap of some q-lamination and is pre-periodic so that the forward orbit of m' does not contain m'', and the forward orbit of m'' does not contain m'. Thus, if $m_1 \leqslant m$ are two minors and m is an eventual image of m_1, then $m_1 < m$.*

PROOF. (1) By the choice of k, we have $i \geqslant k - 1$. Also, $\sigma_2^k(\ell) = m$ implies that $\sigma_2^{k-1}(\ell)$ is a horizontal edge of $Q(m)$. If, for some $i > k - 1$, the leaf $\sigma_2^i(\ell)$ is a horizontal edge of $Q(m)$, then m is a periodic minor whose orbit includes a horizontal edge of $Q(m)$. However, the orbit of a periodic minor m includes a major of $\mathcal{L}(m)$ but does not include horizontal edges of $Q(m)$.

(2) Easily follows from the No Wandering Triangles Theorem. □

The next lemma is based on Proposition 2.3.

LEMMA 2.5. *Let m be a minor. Let $a \in H(m)$ be a point and n be the smallest integer such that $\sigma_2^n(a)$ is an endpoint of m. Let ℓ be a leaf of $\mathcal{L}(m)$ with endpoint a chosen so that $\sigma_2^{n-1}(\ell)$ is a major of $\mathcal{L}(m)$. Among all iterated images of ℓ that separate a from m, choose the one closest to m; call it ℓ'. If no iterated image of ℓ separates a from m, set $\ell' = \ell$. Then ℓ' is a from m derived minor.*

The leaf ℓ' is well defined as there are only finitely many iterated images $\ell'' \leqslant m$ of ℓ (this is because no iterated image of m is behind m, which follows from the Central Strip Lemma). Observe that ℓ defined in the lemma never maps to a horizontal edge of $Q(m)$ because $\sigma_2^{n-1}(\ell)$ is a major of $\mathcal{L}(m)$, and majors of $\mathcal{L}(m)$ do not map to horizontal edges of $Q(m)$.

PROOF. By the choice of ℓ the leaf ℓ' is a pullback of m in $\mathcal{L}(m)$ such that no forward image of ℓ' separates m from ℓ' and no image of ℓ' is a horizontal edge of $Q(m)$. Hence by definition ℓ' is a from m derived minor. □

We are ready to prove Theorem A. Observe that by Theorem A a minor $\tilde{m} < m$ is an offspring of a minor m iff $\sigma_2^n(\tilde{m}) = m$ for some $n > 0$.

PROOF OF THEOREM A. Let m be a minor. Let $a \in H(m)$ be a point and n be a minimal integer such that $\sigma_2^n(a)$ is an endpoint of m. Let us find the leaf ℓ' as in Lemma 2.5. Then $\ell' \in \mathcal{L}(m)$ is a from m derived minor which is a child of m. If a is an endpoint of ℓ', we are done. Otherwise we apply Lemma 2.5 to a and ℓ'. Observe that this time we will find the appropriate pullback of ℓ' with endpoint a in the lamination $\mathcal{L}(\ell')$, not in $\mathcal{L}(m)$, and our choice will be made to make sure that this pullback of ℓ' does not pass through a horizontal edge of $Q(\ell')$. On the other hand, the pullback of ℓ' that we will find does eventually map to m. After finitely many steps the just described process will end, and we will find the desired offspring of m with endpoint a. □

We complete this section with two lemmas that will be used later on.

LEMMA 2.6. *Let m be the minor of a lamination $\mathcal{L}$. Then any leaf $\ell \in \mathcal{L}$ such that $\ell \leqslant m$ and $|\ell| > \frac{|m|}{2}$ is a minor. In particular, if $\ell \leqslant m$ is sufficiently close to m, then ℓ is a minor.*

PROOF. By Lemma 1.1, the chord ℓ is minor-like so that SA1 holds for m. Let us verify property SA2 for ℓ. Let $|m| = 2\lambda$. Then the width of the strip $S(m)$ is λ. If $\ell \in \mathcal{L}$, $\ell \leqslant m$ and $|\ell| > \frac{|m|}{2} = \lambda$, then, by Lemma 1.6(1), we have $|\sigma_2^i(\ell)| > \lambda$ for every $i > 0$. Hence eventual images of ℓ do not enter the interior of $S(m)$ horizontally. On the other hand, they cannot enter the interior of $S(m)$ vertically since the edges $\pm M$ of $S(m)$ are the majors of $\mathcal{L}$. Since $\ell \in \mathcal{L}$, eventual images of ℓ do not cross the majors $\pm M$ of $\mathcal{L}$. Hence they do not intersect $S(\ell)$ at all, and ℓ has property SA2. By Lemma 1.6, the leaf ℓ also has property SA3. Hence ℓ is a stand alone minor. □

Lemma 2.7 describes other cases when a minor can be discovered; assumptions of Lemma 2.7 reverse those of Proposition 2.3.

LEMMA 2.7. *Let m be the minor of a lamination $\mathcal{L}$ and $\ell \in \mathcal{L}$ is a minor-like leaf such that $m \leqslant \ell$. Moreover, suppose that $m \leqslant \sigma_2^n(\ell) \leqslant \ell$ is false for any $n > 0$. Then ℓ is a minor. In particular, this is the case if $m \leqslant \ell \leqslant \widehat{m}$ where $\widehat{m} \in \mathcal{L}$ is a minor, $\sigma_2^n(\ell) = \widehat{m}$ for some n, and no leaf $\sigma_2^i(\ell)$, $0 < i < n$, separates m from ℓ.*

PROOF. By the assumptions, SA1 holds for ℓ. By Lemma 1.6(2), property SA2 also holds for ℓ. To verify SA3, assume, by way of contradiction, that for some minimal $n > 0$ we have $|\sigma_2^n(\ell)| < |\ell|$. Then by the Central Strip Lemma (which applies because of SA2), the leaf $\sigma_2^{n-1}(\ell) \subset S(\ell)$ is vertical. The fact that m is the minor of $\mathcal{L}$ now implies that $\sigma_2^{n-1}(\ell)$ must be a vertical leaf in $\overline{S(\ell) \setminus S(m)}$ which in turn implies that $m \leqslant \sigma_2^n(\ell) \leqslant \ell$, a contradiction. Thus, SA3 holds for ℓ, and ℓ is a minor.

To prove the second claim of the lemma notice that by the Central Strip Lemma, no eventual image of $\widehat{m}$ is behind $\widehat{m}$. Together with the assumptions of the lemma on ℓ it implies that no eventual image of ℓ separates ℓ from $\widehat{m}$. By the above, ℓ is a minor. □

3. Coexistence and tuning

We start with a general property of minor leaf laminations. A chord ℓ is said to *coexist* with a lamination $\mathcal{L}$ if no leaf of $\mathcal{L}$ is linked with ℓ.

LEMMA 3.1. *Let m be a minor, and $\mathcal{L}(m)$ the corresponding minor leaf lamination. If $Q \subset S(m)$ is a collapsing quadrilateral whose vertical edges coexist with $\mathcal{L}(m)$, then Q is contained in the critical gap of $\mathcal{L}(m)$.*

PROOF. If a horizonal edge ℓ_h of Q and a leaf $\ell \in \mathcal{L}(m)$ cross in $\mathbb{D}$, then, since ℓ cannot cross the vertical edges of Q, ℓ must cross $-\ell_h$. Thus, ℓ is a vertical leaf of $\mathcal{L}(m)$ in $S(m)$, a contradiction. Hence horizonal edges of Q also coexist with $\mathcal{L}(m)$. Since m is non-degenerate, $\mathcal{L}(m)$ has no critical leaves. Thus Q is contained in the critical gap of $\mathcal{L}(m)$. □

Coexistence of chords turns out to be stable under σ_2.

LEMMA 3.2. *Suppose that a chord ℓ coexists with a quadratic lamination $\mathcal{L}$. Then $\sigma_2(\ell)$ also coexists with $\mathcal{L}$.*

PROOF. Assume the contrary: $\sigma_2(\ell)$ is linked with some leaf ab of $\mathcal{L}$. The chords $\pm\ell$ divide the circle $\mathbb{S}$ into four arcs, which will be called the $\pm\ell$-arcs. The two σ_2-preimages of a are in the opposite (=not adjacent) $\pm\ell$-arcs. Similarly, the two preimages of b are in the remaining opposite $\pm\ell$-arcs. It follows that any pullback of ab in $\mathcal{L}$ crosses ℓ or $-\ell$, a contradiction. □

Two laminations $\mathcal{L}_1$, $\mathcal{L}_2$ are said to *coexist* if no leaves $\ell_1 \in \mathcal{L}_1$ and $\ell_2 \in \mathcal{L}_2$ cross. Thus, coexistence of quadratic laminations is a symmetric relation.

LEMMA 3.3. *Let m_1 be a minor that is an offspring of a non-degenerate minor m_0. If $\mathcal{L}(m_1)$ coexists with some quadratic lamination $\mathcal{L} \neq \mathcal{L}(m_1)$ with minor m, then either m is an endpoint of m_1 and $\mathcal{L}$ is the corresponding lamination with a critical leaf, or $\mathcal{L}$ is the q-lamination associated to $\mathcal{L}(m_1)$, or $m_1 < m_0 \leqslant m$.*

PROOF. We assume from the very beginning that m is not an endpoint of m_1. It is easy to see that m_1 is non-periodic since m_1 is an offspring of m_0. Hence by Theorem 1.7 the lamination $\mathcal{L}(m_1)$ contains the critical quadrilateral $Q(m_1)$, and $\mathcal{L}(m_1)$ is obtained from the q-lamination $\widehat{\mathcal{L}}(m_1)$ (with finite gaps and all leaves being non-isolated) by inserting vertical edges of $Q(m_1)$ in its central gap $C(\widehat{\mathcal{L}}(m_1))$ (in this way one adds $Q(m_1)$ to $\widehat{\mathcal{L}}(m_1)$) and then pulling them back within $\widehat{\mathcal{L}}(m_1)$. The only two laminations that tune $\mathcal{L}(m_1)$ are the ones whose minors are endpoints of m_1. Hence, by our assumption, it follows that $\mathcal{L}$ cannot have any leaves that do not belong to $\mathcal{L}(m_1)$. In other words, $\mathcal{L} \subsetneqq \mathcal{L}(m_1)$. Since the majors $\pm M$ of $\mathcal{L}$ are leaves of $\mathcal{L}(m_1)$, then they are located so that $S(m) \supset S(m_1)$ and hence $m_1 \leqslant m$.

Consider the case when $m_1 \in \mathcal{L}$. If the majors $\pm M_1$ belong to $\mathcal{L}$, it follows that $m = m_1$. Since $m = m_1$ is not periodic, the central gap of $\mathcal{L}$ must be finite. Since by

Theorem 1.7 the horizontal edges of $Q(m_1)$ are limits of leaves of $\widehat{\mathcal{L}}(m_1) \subset \mathcal{L}(m_1)$, $Q(m_1)$ must be a gap of $\mathcal{L}$, and it follows that $\mathcal{L} = \mathcal{L}(m_1)$. Suppose that $m_1 \in \mathcal{L}$ but $\pm M_1$ do not belong to $\mathcal{L}$. Let $\widehat{C}$ be the critical set of the q-lamination $\widehat{\mathcal{L}}$ associated to m_1. Then the horizontal edges of $Q(m_1)$ must be edges of $\widehat{C}$ and leaves of $\mathcal{L}$. Indeed, some leaves of $\mathcal{L}$ must map to m_1, and the vertical edges of $Q(m_1)$ do not belong to $\mathcal{L}$. Hence the critical set of $\mathcal{L}$ is a gap H containing the horizontal edges of $Q(m_1)$ in the boundary. Consider two cases.

If H is finite, then the fact that $\pm M_1$ do not belong to $\mathcal{L}$ and the fact that edges of $\widehat{C}$ are approached from the outside of $\widehat{C}$ by leaves of $\widehat{\mathcal{L}} \subset \mathcal{L}(m_1)$ imply that $H \subset \widehat{C}$ is different from $Q(m_1)$. Since no edges of H can cross $\pm M_1$ and the images of the edges of H must be edges of $\sigma_2(\widehat{C})$ (otherwise some of their eventual images will cross), we have $H = \widehat{C}$ and, hence, $\mathcal{L} = \widehat{\mathcal{L}}$ is the q-lamination associated to m_1.

If H is infinite, then H is a quadratic Fatou gap, and m_1 is an edge of its image; it is well known that then H is periodic of period, say, n. It is known that there is a unique periodic edge M of H, and it is of period n. Moreover, M and its sibling $-M$ are the majors of the unique lamination that has H as its gap; this lamination is in fact a q-lamination and, evidently, it has to coincide with $\mathcal{L}$ so that $m = \sigma_2(M)$ is an edge (actually, unique periodic edge) of $\sigma_2(H)$. It is known that all edges of H eventually map to m (it is a consequence of the Central Strip Lemma), in particular so does m_1 (which is an edge of $\sigma_2(H)$) and m_0 (which is an eventual image of m_1).

The Central Strip Lemma also imposes restrictions on possible locations of iterated images of H. Namely, the entire gap $\sigma_2(H)$ is located under m while all other iterated images of H are located on the other side of m. Now, m_0 is a minor of some lamination and an eventual image of m_1. Since m is an eventual image of m_1, it follows that m is an eventual image of m_0. If m_0 is an edge of some iterated image of H different from $\sigma_2(H)$, then $m_1 \leqslant m_0$ implies $m \leqslant m_0$ (recall that both m_1 and m are edges of $\sigma_2(H)$). Since m_0 is eventually mapped to $m \leqslant m_0$, we must have $m = m_0$, and we are done in this case. Thus we may assume that m_0 is an edge of $\sigma_2(H)$. Since the only edge ℓ of H such that $m_1 < \ell$ is the edge m, the fact that $m_1 < m_0$ implies again that $m_0 = m$. All that covers the "trivial" cases included in the theorem.

Now, if $m_1 \notin \mathcal{L}$, then m_1 is a diagonal of a gap G of $\mathcal{L}$ whose edges are leaves of $\mathcal{L}(m_1)$. Since m_1 is approached by uncountably many leaves of $\mathcal{L}(m_1)$ from at least one side, $G \cap \mathbb{S}$ is infinite and uncountable (in particular, G is not an iterated pullback of a caterpillar gap). Also, G is not an iterated pullback of a periodic Siegel gap as otherwise m_1, being a diagonal of G, will have some eventual images that cross. Since G is infinite, it is eventually precritical and an image $\sigma_2^i(G) = H$ of G is a periodic critical quadratic Fatou gap containing as a diagonal the leaf $\sigma_2^i(m_1)$. As in the previous paragraph, there is a unique periodic edge M of H, and it is of period n. Moreover, M and its sibling $-M$ are the majors of a unique lamination that has H as its gap; this lamination is in fact a q-lamination and, evidently, it coincides with $\mathcal{L}$ so that $m = \sigma_2(M)$.

The majors $\pm M_1$ coexist with $\mathcal{L}$ and cannot cross edges of H. Hence $m_1 = \sigma_2(M_1)$ is a diagonal or an edge of $\sigma_2(H)$. Since $m_1 \leqslant m$ and $m_1 \leqslant m_0$, we have that either $m_0 \leqslant m$, or $m \leqslant m_0, m \neq m_0$. By way of contradiction assume that $m \leqslant m_0, m \neq m_0$. However, then under some iteration of σ_2 the leaf m_0, which is an eventual image of m_1, will be mapped back to $\sigma_2(H)$ so that for the appropriate eventual image $\sigma_2^j(m_0)$ of m_0 we have $\sigma_2^j(m_0) \leqslant m \leqslant m_0$, which is only possible for the minor m_0 if in fact $m_0 = m$, a contradiction. Thus, $m_0 \leqslant m$, as desired. □

The next theorem describes some cases when one lamination tunes another one. Recall that, by Definition 1.8, the central gap of a lamination $\mathcal{L}(m_0)$ is either a collapsing quadrilateral or an infinite gap.

THEOREM 3.4. *Given minors m_0, m_1 and m, the following statements hold.*

(1) *If $\mathcal{L}(m_1)$ has majors $\pm M_1$ contained in the central gap of $\mathcal{L}(m_0)$, then $\mathcal{L}(m_0) \subset \mathcal{L}(m_1)$; if $m_1 \neq m_0$, then $\mathcal{L}(m_1) \neq \mathcal{L}(m_0)$.*
(2) *If m, m_0 and m_1 are non-degenerate minors such that m_1 is a child of m_0, the lamination $\mathcal{L}(m_1)$ coexists with $\mathcal{L}(m)$, and m is neither m_1 nor an endpoint of m_1, then $\mathcal{L}(m) \subset \mathcal{L}(m_0)$.*
(3) *If m_1 is an offspring of m_0 and $\mathcal{L}(m) \subsetneqq \mathcal{L}(m_1)$, then $\mathcal{L}(m) \subsetneqq \mathcal{L}(m_0)$.*

PROOF. (1) Let the central gap $C(m_0)$ of $\mathcal{L}(m_0)$ be a collapsing quadrilateral. Then the fact that $\pm M_1 \subset C(m_0)$ implies that $m_1 = m_0$ and $\mathcal{L}(m_1) = \mathcal{L}(m_0)$.

Let now $C(m_0)$ be an infinite gap. Then m_0 is periodic of the same period as $C(m_0)$. Let us write M_0 for the pullback of m_0 that is invariant under the first return map f of $C(m_0) \cap \mathbb{S}$. Assume that M_1 separates $-M_1$ from M_0 (or $M_1 = -M_1$ is critical). Consider iterated pullbacks of M_1 chosen so that each next pullback separates the previous pullback from M_0. By definition of m_1-pullbacks, all these pullbacks belong to $\mathcal{L}(m_1)$. Since these f-pullbacks converge to M_0, we have $M_0 \in \mathcal{L}(m_1)$. Similarly, all edges of $C(m_0)$ are in fact m_1-pullbacks of M_0, which implies that all edges of $C(m_0)$ belong to $\mathcal{L}(m_1)$. In the same way, it follows from definition of m_1-pullbacks that all other leaves of $\mathcal{L}(m_0)$ are in fact leaves of $\mathcal{L}(m_1)$. Hence, $\mathcal{L}(m_0) \subset \mathcal{L}(m_1)$.

(2) Let $\pm M$, $\pm M_i$ be the majors of $\mathcal{L}(m)$, $\mathcal{L}(m_i)$, for $i = 0, 1$. Since the "trivial" cases of Lemma 3.3 do not hold, then by Lemma 3.3 we see that $m_1 < m_0 \leqslant m$. Thus, $S(m_1) \subset S(m_0) \subset S(m)$. Since m_1 maps to either M_0 or $-M_0$ under some iterate of σ_2 (see Proposition 2.3), the majors $\pm M_0$ coexist with $\mathcal{L}(m)$. We have $\pm M_0 \subset S(m_0) \subset S(m)$, therefore, $\pm M_0$ are contained in the central gap $C(m)$ of $\mathcal{L}(m)$. The result now follows from (1).

(3) By Theorem A we may assume that $m_1 < m_{(n-1)/n} < \cdots < m_{1/n} < m_0$ where $m_{(i+1)/n}$ is a child of $m_{i/n}$ for $i = 0, \ldots, n-1$. Applying (2) inductively, we see that $\mathcal{L}(m) \subsetneqq \mathcal{L}(m_{(n-1)/2}), \ldots, \mathcal{L}(m) \subsetneqq \mathcal{L}(m_0)$. □

4. Almost non-renormalizable minors: proof of Theorem B

We begin by discussing which minors can be approximated by offsprings of a given minor. Recall the following fact.

LEMMA 4.1 ([**Thu85**], Lemma II.6.10a, part (b)). *Let m_0 be a non-degenerate minor. If $m \leqslant m_0$ is a minor, then $m_0 \in \mathcal{L}(m)$. In particular, $\sigma_2^n(m)$ cannot cross m_0 for $n > 0$.*

The next lemma elaborates on Lemma 2.5.

LEMMA 4.2. *Suppose that $\tilde{m} < m$ are two minors and $\sigma_2^n(\tilde{m}) = m$ for a minimal $n > 0$. Then the following holds.*

(1) *If no image $\sigma_2^i(\tilde{m})$ for $0 < i < n$ is a minor separating $\tilde{m}$ from m, then $\tilde{m}$ is a child of m (in particular, $\tilde{m} \in \mathcal{L}(m)$).*
(2) *Let $\tilde{m} = m_0 < m_1 < \cdots < m_{r-1} < m_r = m$ be all images of $\tilde{m}$ that are minors separating $\tilde{m}$ from m. Then m_i is a child of m_{i+1} for $0 \leqslant i \leqslant r-1$.*

PROOF. (1) To prove that $\tilde{m} \in \mathcal{L}(m)$, consider $\sigma_2^i(\tilde{m})$ for $0 \leqslant i \leqslant n-1$. Choose the greatest $i < n$ such that $\sigma_2^i(\tilde{m}) = m'$ satisfies $\tilde{m} \leqslant m' \leqslant m$. Then no iterated image of m' separates $\tilde{m}$ from m. We claim that no image of m' enters $S(m)$ vertically. Indeed, otherwise the next image of m' would have to enter $C(m)$ either separating $\tilde{m}$ and m (impossible by the choice of i), or behind $\tilde{m}$ (impossible because $\tilde{m}$ is a minor). Hence the leaves $\sigma_2^j(\tilde{m})$, where $j = n-1, n-2, \ldots, i$ are pullbacks of m in $\mathcal{L}(m)$. Thus, $m' \in \mathcal{L}(m)$ and is, therefore, a from m derived minor. If $i > 0$, then m' is a minor separating $\tilde{m}$ from m, a contradiction with the assumptions of the lemma. We must conclude that $i = 0$ and $m' = \tilde{m}$, in particular, $\tilde{m} \in \mathcal{L}(m)$. By definition, it follows that $\tilde{m}$ is a child of m.

(2) Follows from (1) applied to pairs of minors $m_i < m_{i+1}$, where $0 \leqslant i \leqslant r-1$. □

The following lemma relates approximation by dynamical pullbacks and approximation by parameter pullbacks.

LEMMA 4.3. *Let m_0 be a non-degenerate minor. Suppose that $m \leqslant m_0$ is a minor approximated by pullbacks of m_0 in $\mathcal{L}(m)$. Then m can be approximated by offsprings of m_0.*

PROOF. We may assume that m is never mapped to m_0 under σ_2. By Lemma 4.1, the chord m_0 is a leaf of $\mathcal{L}(m)$. Let ℓ_n be a sequence of leaves of $\mathcal{L}(m)$ converging to m and such that $\sigma_2^{k_n}(\ell_n) = m_0$ for some k_n. Since infinitely many ℓ_n's cannot share an endpoint with m, then we may assume that all ℓ_n are disjoint from m in $\overline{\mathbb{D}}$. We may assume that $\ell_n < m_0$. If $\ell_n < m$ for infinitely many values of n, then, by Lemma 2.6, we may assume that these ℓ_n are minors, and, by Lemma 4.2 and Theorem A, they are offsprings of m_0. Suppose now that $\ell_n > m$ for infinitely many values of n; we may assume this is true for all n. Consider all images of ℓ_n that separate m from m_0 and choose among them the closest to m leaf $\sigma_2^i(\ell)$. By Lemma 2.7 $\sigma_2^i(\ell)$ is a minor, and by Theorem A $\sigma_2^i(\ell)$ is an offspring of m_0. This completes the proof of the lemma. □

We can now prove the following theorem.

THEOREM 4.4. *Let m_0 be a periodic non-degenerate minor, and let $m \leqslant m_0$ be a non-degenerate minor. Suppose that any lamination $\mathcal{L} \subsetneqq \mathcal{L}(m)$ satisfies $\mathcal{L} \subsetneqq \mathcal{L}(m_0)$. Then m is a limit of offsprings of m_0.*

PROOF. By Lemma 4.3, it suffices to approximate m by pullbacks of m_0 in $\mathcal{L}(m)$. Consider the lamination $\mathcal{L}_1$ consisting of iterated pullbacks of m_0 in $\mathcal{L}(m)$ and their limits (this includes the iterated images of m_0 since m_0 is periodic); then $\mathcal{L}_1 \subset \mathcal{L}(m)$. If $\mathcal{L}_1 = \mathcal{L}(m)$, we are done; let $\mathcal{L}_1 \neq \mathcal{L}(m)$. Then, by our assumption, $\mathcal{L}_1 \subsetneqq \mathcal{L}(m_0)$. However, since $m_0 \in \mathcal{L}_1$, it follows from Theorem 1.7 that $\mathcal{L}_1 = \mathcal{L}(m_0)$, a contradiction with our assumption. □

We need a lemma dealing with tuning of q-laminations.

LEMMA 4.5. *Let $\mathcal{L}_1 \subsetneqq \mathcal{L}_2$ be q-laminations where $\mathcal{L}_1$ is not the empty lamination. Then $\mathcal{L}_1$ has a periodic quadratic Fatou gap, and, therefore, its minor is periodic and non-degenerate.*

PROOF. Suppose that $\mathcal{L}_1$ does not have a periodic quadratic Fatou gap. Then all gaps of $\mathcal{L}_1$ are either (a) finite, or (b) infinite eventually mapped to a periodic *Siegel* gap for whom the first return map is semiconjugate to an irrational rotation (the semiconjugacy collapses the edges of the gap). Evidently, no leaves of $\mathcal{L}_2$ can be contained in finite gaps of $\mathcal{L}_1$ because both laminations are q-laminations. On the other hand, no leaves of $\mathcal{L}_2$ can be contained in periodic Siegel gaps because any such leaf would cross itself under a suitable power of σ_2 (this conclusion easily follows from the semiconjugacy with an irrational rotation). Thus, if $\mathcal{L}_1$ does not have a periodic quadratic Fatou gap then no new leaves can be added to $\mathcal{L}_1$ and the inclusion $\mathcal{L}_1 \subsetneqq \mathcal{L}_2$ is impossible. □

Recall that a quadratic lamination $\mathcal{L}$ is called *almost non-renormalizable* if $\mathcal{L}' \subsetneqq \mathcal{L}$ implies that $\mathcal{L}'$ is the empty lamination. Note that all almost non-renormalizable laminations with non-degenerate minors are q-laminations (if $\mathcal{L}$ is not a q-lamination with a non-degenerate minor then by Theorem 1.7 there exists a unique non-empty q-lamination $\widehat{\mathcal{L}} \subsetneqq \mathcal{L}$, a contradiction). The role of almost non-renormalizable minors is clear from the next lemma.

LEMMA 4.6. *Let $\mathcal{L}$ be a lamination with non-degenerate minor m. Then there exists a unique almost non-renormalizable lamination $\mathcal{L}_0 \subset \mathcal{L}$ with non-degenerate minor m_0 such that $m \subset \sigma_2(C(m_0))$.*

PROOF. Consider a lamination $\mathcal{L}' \subset \mathcal{L}$ with minor m'. Then, by definition, $m \leqslant m'$. Hence minors of all laminations contained in $\mathcal{L}$ are linearly ordered. Take the intersection $\mathcal{L}_0$ of all non-empty laminations contained in $\mathcal{L}$; note that this intersection is not the empty lamination as every non-empty lamination contains a leaf of length at least $\frac{1}{3}$. It follows that $\mathcal{L}_0$ is itself a non-empty lamination and that the minor m_0 of $\mathcal{L}_0$ is such that $m' \leqslant m_0$ for every non-empty lamination $\mathcal{L}' \subset \mathcal{L}$ (here m' is the minor of $\mathcal{L}'$). Evidently, $m \subset \sigma_2(C(m_0))$ (notice that if $\mathcal{L}' \subset \mathcal{L}$ then $C(\mathcal{L}') \supset C(\mathcal{L})$). □

The set QML^{nr} by definition consists of all singletons in $\mathbb{S}$ and the postcritical sets of all almost non-renormalizable laminations. The following theorem was obtained [**BOT17**]; for completeness, we prove it below.

THEOREM 4.7. *The set QML^{nr} is a lamination.*

PROOF. We only need to prove that QML^{nr} is closed in the Hausdorff metric. We claim that QML^{nr} is obtained from QML by removing all minors that are contained in the interiors of the gaps $\sigma_2(C(m))$ (except for their endpoints), where m are non-degenerate almost non-renormalizable periodic minors. The theorem will follow from this claim (indeed, the set of removed leaves is open in the Hausdorff metric).

Firstly, we show that a leaf ℓ of QML^{nr} cannot intersect the interior of a gap $G = \sigma_2(C(m))$ with $m \in \mathrm{QML}$. Indeed, otherwise the fact that all our leaves are leaves of QML implies that $\ell \subset G$. Hence the majors $\pm L$ of $\mathcal{L}(\ell)$ are contained in

$C(m)$. By Theorem 3.4, part (1), we have then $\mathcal{L}(m) \subsetneqq \mathcal{L}(\ell)$. By definition, this contradicts the fact that ℓ is a minor of an almost non-renormalizable lamination.

Secondly, suppose that $\tilde{m}$ is a minor that does not intersect the interior of any gap $\sigma_2(C(m))$, where m is a non-degenerate periodic almost non-renormalizable minor. We may assume that $\tilde{m}$ is non-degenerate. We claim that $\tilde{m} \in \mathrm{QML}^{nr}$, i.e. that $\tilde{m}$ is an edge of the postcritical set of an almost non-renormalizable lamination. By way of contradiction, assume otherwise. Observe that $\tilde{m}$ is an edge of the postcritical set of the q-lamination $\widehat{\mathcal{L}}(\tilde{m})$. By the assumption, it follows that $\widehat{\mathcal{L}}(\tilde{m})$ is not almost non-renormalizable. Hence by Lemmas 4.5 and 4.6 there exists a non-empty almost non-renormalizable lamination $\mathcal{L}'$ such that $\tilde{m} \subset \sigma_2(C(\mathcal{L}'))$, a contradiction with the assumption on $\tilde{m}$.

□

Let m_0 be a non-degenerate periodic minor. Define the set $\mathrm{OL}(m_0)$ consisting of all offsprings of m_0 and their limits. The following theorem is a reformulation of Theorem B.

THEOREM 4.8. *The lamination* QML^{nr} *is the union of* $\mathrm{OL}(m_0)$*, where* m_0 *runs through all edges of* CA^c*.*

PROOF. Consider an almost non-renormalizable minor $m \in \mathrm{QML}^{nr}$. There is an edge m_0 of the combinatorial main cardioid such that $m \leqslant m_0$. We claim that $m \in \mathrm{OL}(m_0)$. Indeed, consider all pullbacks of m_0 in $\mathcal{L}(m)$ and all limit leaves of such pullbacks. By [**BMOV13**], this collection $\mathcal{L}'$ of leaves is a lamination, and by construction $\mathcal{L}' \subset \mathcal{L}(m)$. Since $\mathcal{L}(m)$ is almost non-renormalizable, $\mathcal{L}' = \mathcal{L}$. Hence, pullbacks of m_0 in $\mathcal{L}(m)$ approximate m. By Lemma 4.3, the minor m is approximated by offsprings of m_0.

Now, let $m \in \mathrm{OL}(m_0)$, where m_0 is an edge of CA^c. Then there is a sequence of minors ℓ_i converging to m such that each ℓ_i is an offspring of m_0. We claim that m is almost non-renormalizable, i.e., that $m \in \mathrm{QML}^{nr}$. Assume the contrary: m is contained in a gap U of QML^{nr} and intersects the interior of U. The only way it can happen is when $U = \sigma_2(C(m_1))$ is the postcritical gap of an almost non-renormalizable lamination $\mathcal{L}(m_1)$. Then ℓ_i must also intersect the interior of U for some i, hence ℓ_i must be contained in U. By Theorem 3.4, part (1), we have $\mathcal{L}(m_1) \subsetneqq \mathcal{L}(\ell_i)$. Since ℓ_i is an offspring of m_0, it follows by Theorem 3.4, part (3), that $\mathcal{L}(m_1) \subsetneqq \mathcal{L}(m_0)$. However, this is impossible because m_0 itself is almost non-renormalizable, and the only lamination strictly contained in $\mathcal{L}(m_0)$ is the empty lamination. □

Acknowledgments. The first named author started discussing combinatorial models for parameter spaces of cubic polynomials with the third named author while visiting Jacobs University in Bremen in 2009. We are pleased to thank Jacobs University for their hospitality.

The authors also would like to thank the referees for valuable remarks and suggestions.

References

[BH01] X. Buff and C. Henriksen, *Julia sets in parameter spaces*, Comm. Math. Phys. **220** (2001), no. 2, 333–375, DOI 10.1007/PL00005568. MR1844629

[BMOV13] Alexander M. Blokh, Debra Mimbs, Lex G. Oversteegen, and Kirsten I. S. Valkenburg, *Laminations in the language of leaves*, Trans. Amer. Math. Soc. **365** (2013), no. 10, 5367–5391, DOI 10.1090/S0002-9947-2013-05809-6. MR3074377

[BO04] Alexander Blokh and Lex Oversteegen, *Wandering triangles exist* (English, with English and French summaries), C. R. Math. Acad. Sci. Paris **339** (2004), no. 5, 365–370, DOI 10.1016/j.crma.2004.06.024. MR2092465

[BO08] Alexander Blokh and Lex Oversteegen, *Wandering gaps for weakly hyperbolic polynomials*, Complex dynamics, A K Peters, Wellesley, MA, 2009, pp. 139–168, DOI 10.1201/b10617-4. MR2508256

[BOPT16] Alexander Blokh, Lex Oversteegen, Ross Ptacek, and Vladlen Timorin, *The combinatorial Mandelbrot set as the quotient of the space of geolaminations*, Dynamics and numbers, Contemp. Math., vol. 669, Amer. Math. Soc., Providence, RI, 2016, pp. 37–62, DOI 10.1090/conm/669/13422. MR3546662

[BOPT16b] Alexander Blokh, Lex Oversteegen, Ross Ptacek, and Vladlen Timorin, *Laminations from the main cubioid*, Discrete Contin. Dyn. Syst. **36** (2016), no. 9, 4665–4702, DOI 10.3934/dcds.2016003. MR3541500

[BOPT16c] Alexander Blokh, Lex Oversteegen, Ross Ptacek, and Vladlen Timorin, *The parameter space of cubic laminations with a fixed critical leaf*, Ergodic Theory Dynam. Systems **37** (2017), no. 8, 2453–2486, DOI 10.1017/etds.2016.22. MR3719267

[BOPT17] A. Blokh, L. Oversteegen, R. Ptacek, V. Timorin, *Laminational models for some spaces of polynomials of arbitrary degree*, preprint arXiv:1401.5123, to appear in Memoirs of the American Mathematical Society (2017)

[BOT17] Alexander Blokh, Lex Oversteegen, and Vladlen Timorin, *Perfect subspaces of quadratic laminations*, Sci. China Math. **61** (2018), no. 12, 2121–2138, DOI 10.1007/s11425-017-9305-3. MR3881952

[DH85] A. Douady and J. H. Hubbard, *Étude dynamique des polynômes complexes. Partie I* (French), Publications Mathématiques d'Orsay [Mathematical Publications of Orsay], vol. 84, Université de Paris-Sud, Département de Mathématiques, Orsay, 1984. MR762431

[EY99] Adam Epstein and Michael Yampolsky, *Geography of the cubic connectedness locus: intertwining surgery* (English, with English and French summaries), Ann. Sci. École Norm. Sup. (4) **32** (1999), no. 2, 151–185, DOI 10.1016/S0012-9593(99)80013-5. MR1681808

[Kel00] Karsten Keller, *Invariant factors, Julia equivalences and the (abstract) Mandelbrot set*, Lecture Notes in Mathematics, vol. 1732, Springer-Verlag, Berlin, 2000. MR1761576

[Lav86] Pierre Lavaurs, *Une description combinatoire de l'involution définie par M sur les rationnels à dénominateur impair* (French, with English summary), C. R. Acad. Sci. Paris Sér. I Math. **303** (1986), no. 4, 143–146. MR853606

[Lav89] P. Lavaurs, *Systèmes dynamiques holomorphes, Explosion de points périodiques paraboliques*, Thèse de doctorat, Université Paris-Sud, Orsay (1989).

[McM00] Curtis T. McMullen, *The Mandelbrot set is universal*, The Mandelbrot set, theme and variations, London Math. Soc. Lecture Note Ser., vol. 274, Cambridge Univ. Press, Cambridge, 2000, pp. 1–17. MR1765082

[Mil00] John Milnor, *Dynamics in one complex variable*, 3rd ed., Annals of Mathematics Studies, vol. 160, Princeton University Press, Princeton, NJ, 2006. MR2193309

[PR08] Carsten L. Petersen and Pascale Roesch, *The Yoccoz combinatorial analytic invariant*, Holomorphic dynamics and renormalization, Fields Inst. Commun., vol. 53, Amer. Math. Soc., Providence, RI, 2008, pp. 145–176. MR2477422

[Sch09] D. Schleicher, *Appendix: Laminations, Julia sets, and the Mandelbrot set*, in: "Complex dynamics: Families and Friends", ed. by D. Schleicher, A K Peters (2009), 111–130.

[Thu85] William P. Thurston, *On the geometry and dynamics of iterated rational maps*, Complex dynamics, A K Peters, Wellesley, MA, 2009, pp. 3–137, DOI 10.1201/b10617-3. Edited by Dierk Schleicher and Nikita Selinger and with an appendix by Schleicher. MR2508255

Department of Mathematics, University of Alabama at Birmingham, Birmingham, 35294
Email address: ablokh@math.uab.edu

Department of Mathematics, University of Alabama at Birmingham, Birmingham, 35294
Email address: overstee@uab.edu

Faculty of Mathematics, HSE University, 6 Usacheva St., 119048 Moscow, Russia
Email address: vtimorin@hse.ru

Contemporary Mathematics
Volume **744**, 2020
https://doi.org/10.1090/conm/744/14987

Multi-sensitivity, multi-transitivity and Δ-transitivity

Piotr Oprocha, Tao Yu, and Guohua Zhang

Abstract. In this paper firstly we construct for each $m \in \mathbb{N} \setminus \{1\}$ a weakly mixing system which is $(1, \cdots, m-1)$-sensitive but not $(1, \cdots, m-1, m)$-sensitive, and a minimal system which is $(1, 2)$-sensitive but not $(1, 2, 3)$-sensitive.

It is known that Δ-$(1, 2)$ transitivity implies weak mixing. We will show that, though (a, b)-transitivity implies total transitivity for all $a, b \in \mathbb{N}$, for a general vector $\mathbf{a} \in \mathbb{N}^r, r \in \mathbb{N}$, Δ-$\mathbf{a}$ transitivity implies weak mixing if and only if the vector $\mathbf{a}$ satisfies certain conditions. Furthermore, we prove that this difference will disappear for measurable systems, that is, for measure-theoretical setting multi-ergodicity is equivalent to weak mixing.

1. Introduction

Throughout this paper by a *topological dynamical system* (*t.d.s.* for short) we mean a pair (X, T), where X is a compact metric space with a metric d and $T\colon X \to X$ is a continuous surjection. Denote the sets of all integers, nonnegative integers and natural numbers by $\mathbb{Z}$, $\mathbb{N}_0$ and $\mathbb{N}$ respectively.

The notion of sensitivity (sensitive dependence on initial conditions) originated from a paper by Ruelle [**20**]. Auslander and Yorke [**4**] started to call a t.d.s. (X, T) *sensitive* if there exists $\delta > 0$ such that for every $x \in X$ and every neighborhood U_x of x, there exist $y \in U_x$ and $n \in \mathbb{N}$ with $d(T^n x, T^n y) > \delta$. The following dichotomy theorem is proved in [**4**]: a minimal system is either equicontinuous or sensitive. Since then, the relationship between sensitivity, transitivity and mixing from topological perspective has been discussed by many researchers.

For a t.d.s. (X, T), $\delta > 0$ and an *opene* (open and nonempty) subset U, denote

$$N_T(U, \delta) = \{n \in \mathbb{N} : \mathrm{diam}(T^n(U)) > \delta\},$$

2010 *Mathematics Subject Classification.* 37B05, 37B20, 37A25.

Key words and phrases. Multi-sensitivity, multi-transitivity, Δ-transitivity, multi-ergodicity.

The first author was supported by National Science Centre, Poland (NCN), grant no. 2015/17/B/ST1/01259 and by the Faculty of Applied Mathematics AGH UST statutory tasks within subsidy of Ministry of Science and Higher Education.

The second author was supported by China Postdoctoral Science Foundation, grant no. 2018M631996.

The third author was supported by NSFC, grants no. 11671094 and 11722103.

where diam($\bullet$) denotes diameter of a subset $\bullet$. It is easy to see that (X,T) is sensitive if and only if there exists $\delta > 0$ such that $N_T(U,\delta) \neq \emptyset$ for each opene subset U. In [**18**] Moothathu introduced the notion of multi-sensitivity. A t.d.s. (X,T) is *multi-sensitive* if there exists $\delta > 0$ such that for any $k \in \mathbb{N}$ and any opene subsets $U_1, U_2, \cdots, U_k \subset X$ we have $\bigcap_{i=1}^{k} N_T(U_i,\delta) \neq \emptyset$. Later Huang, Kolyada and Zhang [**13**] proved that a minimal system is either multi-sensitive or an almost one-to-one extension of its maximal equicontinuous factor. Inspired by [**11**–**13**, **18**], Yu [**22**] introduced the notion of l-sensitivity and constructed a minimal system which is l-sensitive but not $(l+1)$-sensitive. This was further generalized by Jiao et al. [**14**], who introduced the notion of multi-sensitivity with respect to a vector. Let (X,T) be a t.d.s. and $\mathbf{a} = (a_1, a_2, \cdots, a_r)$ be a vector in $\mathbb{N}^r, r \in \mathbb{N}$. A t.d.s. (X,T) is $\mathbf{a}$-*sensitive* if there exists $\delta > 0$ such that $\bigcap_{i=1}^{r} N_{T^{a_i}}(U_i,\delta) \neq \emptyset$ for any opene sets $U_1, \cdots, U_r$. Any $(1,2,3)$-sensitive system is clearly $(1,2)$-sensitive. We will show that the converse is not true by constructing a minimal system which is $(1,2)$-sensitive but not $(1,2,3)$-sensitive (see Example 3.8). Due to technical difficulties, it seems not easy to construct such a system for general $(1,2,\cdots,m)$-sensitivity, however for each $m \in \mathbb{N} \setminus \{1\}$ we were able to construct a weakly mixing system which is $(1,\cdots,m-1)$-sensitive but not $(1,\cdots,m-1,m)$-sensitive (see Example 3.7). These examples are not minimal. In fact, we show that any minimal weakly mixing system is strongly multi-sensitive, that is, $\mathbf{a}$-sensitive for any vector $\mathbf{a} \in \mathbb{N}^r, r \in \mathbb{N}$ (see Corollary 3.10).

The notions of multi-transitivity and Δ-transitivity were introduced by Moothathu in [**19**]. He showed that for a general system Δ-$(1,2)$ transitivity is strictly stronger than weak mixing (cf. the proof of [**19**, Proposition 3]). Then in [**7**, **8**] Chen, Li and Lü systematically studied multi-transitivity and Δ-transitivity, and characterized them by properties of the hitting time sets. We prove that (a,b)-transitivity implies total transitivity for all $a,b \in \mathbb{N}$ (see Theorem 3.4). For any vector $\mathbf{a} = (a_1, \cdots, a_r) \in \mathbb{N}^r$ with $a_1 < \cdots < a_r$ and $r \geq 2$, we show that Δ-$\mathbf{a}$ transitivity implies weak mixing if and only if the vector $\mathbf{a}$ satisfies certain conditions (for details see Theorem 4.2 and Theorem 4.4). As a direct corollary of it, we also obtain that Δ-$(1,2)$ transitivity implies weak mixing.

In [**16**] Kwietniak and Oprocha showed that in general there is no connection between multi-transitivity and weak mixing, and Moothathu proved that for a minimal system multi-transitivity is equivalent to weak mixing [**19**]. It is natural to introduce for measurable dynamical systems a notion similar to multi-transitivity, which we call *multi-ergodicity.* We find that it satisfies a measure-theoretic analogue of Moothathu's result, that is, multi-ergodicity is equivalent to weak mixing for measurable systems.

The structure of the paper is the following. In Section 2, we recall some basic notions which will be used later in the paper. In Section 3, we construct for each $m \in \mathbb{N} \setminus \{1\}$ a weakly mixing system which is $(1,\cdots,m-1)$-sensitive but not $(1,\cdots,m-1,m)$-sensitive, and a minimal system which is $(1,2)$-sensitive but not $(1,2,3)$-sensitive. We also show that (a,b)-transitivity implies total transitivity for all $a,b \in \mathbb{N}$. In Section 4, given a vector $\mathbf{a} = (a_1, \cdots, a_r) \in \mathbb{N}^r$ with $a_1 < \cdots < a_r$ and $r \geq 2$, we provide a criterion for that Δ-$\mathbf{a}$ transitivity implies weak mixing. In Section 5, we prove that multi-ergodicity is equivalent to weak mixing for measurable dynamical systems.

2. Preliminaries

In this section we recall some basic concepts and notions which will be used later.

2.1. Topological dynamical systems. A t.d.s. (X, T) is *transitive* if for all opene subsets U and V, $N_T(U, V) = \{n \in \mathbb{N}_0 : U \cap T^{-n}V \neq \emptyset\}$ is infinite; is *totally transitive* if (X, T^n) is transitive for each $n \in \mathbb{N}$; and is *weakly mixing* if $(X \times X, T \times T)$ is transitive. We say that $x \in X$ is a *transitive point* if its *orbit* $\mathrm{Orb}(x, T) = \{x, Tx, T^2x, \cdots\}$ is dense in X. It is well known that if (X, T) is transitive then the set of all transitive points forms a dense G_δ subset of X. A t.d.s. (X, T) is *minimal* if each point in X is a transitive point.

Let (X, T) and (Y, S) be two t.d.s. If there is a continuous surjection $\pi : X \to Y$ with $\pi \circ T = S \circ \pi$, then we say that π is a *factor map* or an *extension*, (Y, S) is a *factor* of (X, T) and (X, T) is an *extension* of (Y, S). The extension $\pi\colon (X, T) \to (Y, S)$ is said to be *almost one-to-one* if $\{x \in X : \pi^{-1}(\pi(x)) = \{x\}\}$ forms a dense G_δ subset of X.

A t.d.s. (X, T) is called *equicontinuous* if for every $\epsilon > 0$ there is $\delta > 0$ such that whenever $x, y \in X$ with $d(x, y) < \delta$ we have $d(T^n x, T^n y) < \epsilon$ for all $n \in \mathbb{N}_0$. It is well known that every t.d.s. (X, T) has a maximal equicontinuous factor (X_{eq}, T_{eq}). See for example [**15**] for more details on equicontinuous systems and these factors.

Let $F \subset \mathbb{N}_0$. We call F *thick* if it contains arbitrarily long blocks of consecutive integers, that is, for every $d \in \mathbb{N}$ there is $n_d \in \mathbb{N}$ such that $\{n_d, n_d+1, \cdots, n_d+d\} \subset F$; call it *syndetic* if it has bounded gaps, that is, there exists $N \in \mathbb{N}$ such that $\{k, k+1, \cdots, k+N\} \cap F \neq \emptyset$ for every $k \in \mathbb{N}$; and call it *thickly syndetic* if for each $d \in \mathbb{N}$ there is a syndetic set $\{w_d^1, w_d^2, \cdots\}$ such that $\{w_d^i, w_d^i + 1, \cdots, w_d^i + d\} \subset F$ for each $i \in \mathbb{N}$. Note that the intersection of finitely many thickly syndetic sets is also thickly syndetic. See for example [**1, 9**] for more such subsets and their relationship with dynamical systems.

2.2. Spacing shift. Let $\Sigma = \{0, 1\}^{\mathbb{N}_0}$ be equipped with the product topology induced by discrete topology on $\{0, 1\}$. A metric defining this topology is given by $d(x, y) = 0$ if $x = y$, and $d(x, y) = 2^{-i}$ if $x \neq y$, where $i = \min\{j : x_j \neq y_j\}$ when $x = x_0x_1x_2\cdots$ and $y = y_0y_1y_2\cdots$. A *word* is a finite sequence of elements of $\{0, 1\}$. For two words $w = x_1 \cdots x_n$ and $v = y_1 \cdots y_m$ define $wv = x_1 \cdots x_n y_1 \cdots y_m$. If w is a word and $n \geq 1$ then by w^n we denote a word which is a concatenation of n copies of w. If $n = 0$ then w^n is the *empty word.* The *length* of a word w is the number of elements of w, and is denoted $|w|$. We say that a word $w = w_1w_2\cdots w_l$ appears in $x = x_0x_1x_2\cdots \in \Sigma$ at position t if $x_{t+j-1} = w_j$ for $j = 1, \cdots, l$. As usual, for any word w we denote by w_t the element of the sequence w standing at position t. Put $Sp(w) = \{|i - j| : w_i = w_j = 1, 1 \leq i < j \leq |w|\}$. Let $P \subset \mathbb{N}$. We say that a word $w = w_1w_2\cdots w_l$ is P*-admissible* if $Sp(w) \subset P$. Let Σ_P be the subset of Σ consisting of all sequences x such that every word which appears in x is P-admissible.

The *shift* transformation $\sigma : \Sigma \to \Sigma$, given by $(\sigma x)_i = x_{i+1}$ for $i = 0, 1, \cdots$, is a continuous surjection. Any nonempty closed subset $X \subset \Sigma$ invariant for σ (i.e., $\sigma(X) = X$) is called a *subshift* of Σ. If X is a subshift, then the *language* of X is the set $\mathcal{L}(X)$ of all words which appear at some position in some element $x \in X$. The set $\mathcal{L}_n(X)$ consists of all elements of $\mathcal{L}(X)$ of length n. Let $P \subset \mathbb{N}$. It is easy to see that Σ_P is a subshift, and $\mathcal{L}(\Sigma_P)$ is the set of all P-admissible words. We will write σ_P for σ restricted to Σ_P, and call the t.d.s. (Σ_P, σ_P) a *spacing shift*. The class of spacing shifts was introduced by Lau and Zame in [**17**], and for a detailed exposition of their properties we refer to [**6**].

Let w be a P-admissible word. By $[w]_P$ we denote the set of all $x \in \Sigma_P$ such that the word w appears at position 0 in x. We call the set $[w]_P$ a *P-admissible cylinder* (a *cylinder* for short). The family of all P-admissible cylinders forms a base of topology of Σ_P inherited from Σ. Note that (Σ_P, σ_P) is weakly mixing if and only if P is a thick set (e.g. see [**6, 17**]).

3. Sensitivity and transitivity with respect to a vector

In this section we study sensitivity and transitivity with respect to a vector.
Firstly we recall the following definitions from [**14, 19**].

Definition 3.1. Let $\mathbf{a} = (a_1, a_2, \cdots, a_r) \in \mathbb{N}^r, r \in \mathbb{N}$. We say that a t.d.s. (X, T) is

(1) *multi-sensitive with respect to* $\mathbf{a}$ (or briefly $\mathbf{a}$-*sensitive*) if there exists $\delta > 0$ such that $\bigcap_{i=1}^{r} N_{T^{a_i}}(U_i, \delta) \neq \emptyset$ for any opene sets $U_1, \cdots, U_r$.
(2) *strongly multi-sensitive* if it is $\mathbf{a}$-sensitive for any vector $\mathbf{a} \in \mathbb{N}^n, n \in \mathbb{N}$.
(3) *multi-transitive with respect to* $\mathbf{a}$ (or briefly $\mathbf{a}$-*transitive*) if $(X^r, T^{(\mathbf{a})})$ is transitive, where $T^{(\mathbf{a})} = T^{a_1} \times T^{a_2} \times \cdots \times T^{a_r}$.
(4) *strongly multi-transitive* if it is $\mathbf{a}$-transitive for any vector $\mathbf{a} \in \mathbb{N}^n, n \in \mathbb{N}$.

Remark 3.2. If (X, T) is $\mathbf{a}$-sensitive, then the above sets $\bigcap_{i=1}^{r} N_{T^{a_i}}(U_i, \delta)$ are infinite.

We also have the following observation.

Lemma 3.3. *Let (X, T) be a transitive system. Then (X, T) is $(1, 2)$-sensitive if and only if there is $\delta > 0$ such that $N_T(U, \delta) \cap N_{T^2}(U, \delta) \neq \emptyset$ for any opene set U.*

Proof. The implication "⇒" is obvious. Now we are going to prove the converse implication "⇐". Observe that for any opene sets U_1, U_2, there is $l \in \mathbb{N}$ such that $U_1 \cap T^{-l}U_2$ is an opene set. Put $V = U_1 \cap T^{-l}U_2$ and take any $n \in N_T(T^{-l}V, \delta) \cap N_{T^2}(T^{-l}V, \delta) \cap \{l+1, l+2, \cdots\}$. Then $\{n-l, 2n-l\} \subset N_T(V, \delta)$ and so $n - l \in N_T(U_1, \delta) \cap N_{T^2}(U_2, \delta)$. □

The following result shows that multi-transitivity, while does not imply weak mixing sometimes, is slightly stronger than transitivity alone.

Theorem 3.4. *Let $a, b \in \mathbb{N}$. If (X, T) is (a, b)-transitive, then (X, T) is totally transitive.*

Proof. Assume the contrary that (X, T^p) is not transitive for some $p \in \mathbb{N}$. Clearly (X, T) is transitive, then we may assume that p is prime by [**5**, Theorem 2.4]. Since (X, T) is (a, b)-transitive, p does not divide a nor b. Applying [**5**, Lemma 2.1 and Theorem 2.3], there exists a regular periodic decomposition

$\{X_0, X_1, \cdots, X_{p-1}\}$ of length p, in particular, $int(X_0), int(X_1), \cdots, int(X_{p-1})$ are pairwise disjoint opene sets and $T^{-l}(int(X_k)) \subset int(X_{k-l(\text{mod } p)})$ for all $0 \leq k \leq p-1$ and $l \geq 0$.

Since (X, T) is (a, b)-transitive, we may choose

$$m \in N_{T^a}(int(X_0), int(X_0)) \cap N_{T^b}(int(X_0), int(X_1)).$$

Then there exists $l_1, l_2 \in \mathbb{N}$ such that $am = pl_1, bm - 1 = pl_2$. Thus p divides m and so $p = 1$, which is a contradiction. Indeed (X, T) is totally transitive. □

It was proved by Moothathu [**19**, Proposition 4] that if (X, T) is a totally transitive t.d.s. with dense periodic points (and so weakly mixing), then (X, T) is strongly multi-transitive. However we construct a system which is transitive with dense periodic points and strongly multi-sensitive, but not totally transitive.

Example 3.5. There is $P \subset \mathbb{N}$ such that the spacing shift (Σ_P, σ_P) is strongly multi-sensitive, but not totally transitive and hence not (a, b)-transitive for all $a, b \in \mathbb{N}$.

Construction of Example 3.5. Put $P = m\mathbb{N}$ for some $m > 1$. Then (Σ_P, σ_P) is transitive and contains dense periodic points by [**6**, Theorem 2.7]. Assume the contrary that it is (a, b)-transitive for some $a, b \in \mathbb{N}$. It is totally transitive by Theorem 3.4, and hence weakly mixing, since it has dense periodic points. But P is not thick, which is a contradiction.

Next we claim that (Σ_P, σ_P) is **a**-sensitive for any vector $\mathbf{a} = (a_1, a_2, \cdots, a_r) \in \mathbb{N}^r, r \in \mathbb{N}$, and hence strongly multi-sensitive. It is enough to consider opene sets $[w_i]_P$, where $w_i \in \mathcal{L}_k(\Sigma_P)$, $i = 1, 2, \cdots, r$. There are integers $0 \leq b_i \leq m-1$ and $n > k$ such that $w_i 0^{a_i n + b_i - k} 1 \in \mathcal{L}(\Sigma_P)$. Clearly $w_i 0^\infty \in [w_i]_P$ therefore $\text{diam}(\sigma^{a_i n}[w_i]_P) \geq \frac{1}{2^m}$ for all $i = 1, 2, \cdots, r$. Indeed (Σ_P, σ_P) is $(a_1, a_2, \cdots, a_r)$-sensitive. □

If a spacing shift (Σ_P, σ_P) is $(a_1, a_2, \cdots, a_r)$-transitive, then for any opene sets $U_1, \cdots, U_r$ there are $k, l \in \mathbb{N}$ with $k < l$ and words $w_i \in \mathcal{L}_k(\Sigma_P)$ such that $[w_i]_P \subset U_i$ and $\sigma_P^{la_i}[w_i]_P \cap [1]_P \neq \emptyset$ for each $i = 1, \cdots, r$. But clearly also $w_i 0^\infty \in U_i$, which proves the following.

Proposition 3.6. *If a spacing shift* (Σ_P, σ_P) *is* $(a_1, a_2, \cdots, a_r)$*-transitive, then* (Σ_P, σ_P) *is* $(a_1, a_2, \cdots, a_r)$*-sensitive.*

The above Proposition gives a clue how to construct a weakly mixing system which is $(1, 2, \cdots, m-1)$-sensitive but not $(1, 2, \cdots, m-1, m)$-sensitive.

Example 3.7. For each $m \in \mathbb{N} \setminus \{1\}$, there is $P \subset \mathbb{N}$ such that the spacing shift (Σ_P, σ_P) is weakly mixing and $(1, 2, \cdots, m-1)$-sensitive but not $(1, 2, \cdots, m-1, m)$-sensitive.

Construction of Example 3.7. Fix each $m \in \mathbb{N} \setminus \{1\}$. Let $P = \bigcup_{k=1}^{\infty} B(k)$, where $B(k) = \{m^{2k-1}, m^{2k-1} + 1, \cdots, m^{2k} - 1\}$. Then $P \subset \mathbb{N}$ is thick, and hence the spacing shift (Σ_P, σ_P) is weakly mixing. Moreover, (Σ_P, σ_P) is $(1, 2, \cdots, m-1)$-transitive by [**16**, Theorem 9], thus is $(1, 2, \cdots, m-1)$-sensitive by Proposition 3.6.

We claim that (Σ_P, σ_P) is not $(1, 2, \cdots, m-1, m)$-sensitive. To prove it, fix any k and let $\delta_k = (\frac{1}{2})^{m^{2k}-1}$. Put $s = m^{2k} - 2$, $w = 10^s 1$ and let $U = [w]_P$. Suppose there is $u \in \mathbb{N}$ large enough such that $\{u, 2u, \cdots, mu\} \subset N_{\sigma_P}(U, \delta_k)$. Then, by the

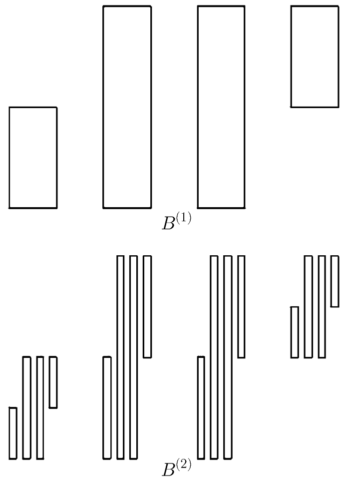


FIGURE 1. First two steps of construction in Example 3.8

definition of P, for each $i = 1, 2, \cdots, m$, there are $k_i \in \mathbb{N}$ and $0 \leq s_i \leq m^{2k} - 1$ satisfying

$$\begin{array}{rcl} m^{2k_i-1} & \leq & iu + s_i \ \leq \ m^{2k_i} - 1, \\ m^{2k_i-1} & \leq & iu + s_i - (m^{2k} - 1) \ \leq \ m^{2k_i} - 1. \end{array}$$

Thus all k_i must be equal. As a consequence, there exists an integer t such that $u \geq m^{2t-1}$ and $mu \leq m^{2t} - 1$ which is impossible. Indeed, the claim holds. □

In the following we construct a minimal system which is $(1, 2)$-sensitive but not $(1, 2, 3)$-sensitive. However, due to technical difficulties, it seems not easy to generalize it for the general case of $(1, 2, \cdots, m)$-sensitivity.

EXAMPLE 3.8. There is a minimal system (X, T) which is $(1, 2)$-sensitive but not $(1, 3)$-sensitive and hence not $(1, 2, 3)$-sensitive.

CONSTRUCTION OF EXAMPLE 3.8. The required system is a modification of the well-known Auslander-Floyd system (for Auslander-Floyd system see for example [**3**, Page 24]).

In what follows by "rectangle" we mean a closed rectangle in the plane with sides parallel to the axes. If B is a rectangle, say $B = [a, a + h] \times [b, b + k]$, then $h(B)$ denotes the union of 4 disjoint rectangles $h(B) = B_0 \cup B_1 \cup B_2 \cup B_3$, where $B_0 = [a, a+\frac{h}{7}]\times[b, b+\frac{k}{2}]$, $B_1 = [a+\frac{2h}{7}, a+\frac{3h}{7}]\times[b, b+k]$, $B_2 = [a+\frac{4h}{7}, a+\frac{5h}{7}]\times[b, b+k]$ and $B_3 = [a+\frac{6h}{7}, a+h]\times[b+\frac{k}{2}, b+k]$. If $K = \bigcup_{i=1}^{n} B_i$ is a disjoint union of rectangles, then we define $h(K) = \bigcup_{i=1}^{n} h(B_i)$.

Now, let $B^{(0)} = [0, 1] \times [0, 1]$, and define inductively $B^{(k+1)} = h(B^{(k)})$ for each $k \in \mathbb{N}_0$. First two steps of this construction are depicted on Figure 1.

Let $h(B^{(0)}) = B_0^{(1)} \cup B_1^{(1)} \cup B_2^{(1)} \cup B_3^{(1)} \subset B^{(0)}$, where rectangles are labeled from left to right. Similarly, $h(B_{a_1}^{(1)}) = B_{a_1,0}^{(2)} \cup B_{a_1,1}^{(2)} \cup B_{a_1,2}^{(2)} \cup B_{a_1,3}^{(2)} \subset B_{a_1}^{(1)}$, where labeling is again from left to right. We extend the construction inductively, putting

$$h(B^{(k)}_{a_1,\cdots,a_k}) = B^{(k+1)}_{a_1,\cdots,a_k,0} \cup B^{(k+1)}_{a_1,\cdots,a_k,1} \cup B^{(k+1)}_{a_1,\cdots,a_k,2} \cup B^{(k+1)}_{a_1,\cdots,a_k,3} \subset B^{(k)}_{a_1,\cdots,a_k},$$

where as before, labeling of rectangles is from left to right.

Let $X = \bigcap_{k=1}^{\infty} B^{(k)}$. Since $B^{(k)}, k \in \mathbb{N}$ is a decreasing sequence of nonempty compact sets, X is a nonempty compact metric space, which consists of vertical line segments, some of which are degenerate (that is, are singletons). A homeomorphism $T : X \to X$ is defined by permuting the rectangles $B^{(k)}_{a_1,\cdots,a_k}$. The permutations of labels are defined by $B^{(k)}_{a_1,\cdots,a_k} \mapsto B^{(k)}_{b_1,\cdots,b_k}$ where $(b_1, \cdots, b_k) = (a_1, \cdots, a_k) + (1, 0, \cdots, 0)$ and the addition is with "carry". These permutations define maps from $\bigcap_{k=1}^{\infty} B^{(k)}_{a_1,\cdots,a_k}$ to $\bigcap_{k=1}^{\infty} B^{(k)}_{b_1,\cdots,b_k}$ provided the intersection is a single point. If these sets are non-degenerate vertical line segments, we map the first of them linearly onto the second. Thus the map T constructed by the above procedure is a homeomorphism permuting the vertical segments.

In fact, (X, T) is a minimal t.d.s. which is an almost one-to-one extension of the 4-adic adding machine. We are going to show that t.d.s. (X, T) has all of the claimed properties.

Step 1. Firstly we prove that (X, T) is $(1, 2)$-sensitive.
Claim 1. For every cylinder $B^{(k)}_{a_1,\cdots,a_k}$ there is $m \in \mathbb{N}$ such that for each $j \in \{1, 2\}$ there are $b_1, \cdots, b_k \in \{1, 2\}$ such that $T^{jm}(B^{(k)}_{a_1,\cdots,a_k}) = B^{(k)}_{b_1,\cdots,b_k}$.

Proof of Claim 1. We define integers m_i for $i = 1, \cdots, k$ in the following way:

$$m_i = \begin{cases} 4^{i-1}, & a_i = 0 \\ 0, & a_i \in \{1, 2\} \\ 4^k - 4^{i-1}, & a_i = 3 \end{cases}.$$

Then for each $j = 1, 2$ and all $c_1, \cdots, c_k \in \{0, 1, 2, 3\}$ we have

$$T^{jm_i}(B^{(k)}_{c_1,\cdots,c_{i-1},a_i,c_{i+1},\cdots,c_k}) = \begin{cases} B^{(k)}_{c_1,\cdots,c_{i-1},j,c_{i+1},\cdots,c_k}, & a_i = 0 \\ B^{(k)}_{c_1,\cdots,c_{i-1},a_i,c_{i+1},\cdots,c_k}, & a_i \in \{1, 2\} \\ B^{(k)}_{c_1,\cdots,c_{i-1},3-j,c_{i+1},\cdots,c_k}, & a_i = 3 \end{cases}.$$

Now put $m = m_1 + \cdots + m_k$. By the above observation we easily see that for each $j = 1, 2$ there are $b_1, \cdots, b_k \in \{1, 2\}$ such that

$$\begin{aligned} T^{jm}(B^{(k)}_{a_1,\cdots,a_k}) &= T^{jm_2+\cdots+jm_k}(T^{jm_1}(B^{(k)}_{a_1,\cdots,a_k})) \\ &= T^{jm_3+\cdots+jm_k}(T^{jm_2}(B^{(k)}_{b_1,a_2,\cdots,a_k})) = \cdots \\ &= T^{jm_k}(B^{(k)}_{b_1,\cdots,b_{k-1},a_k}) = B^{(k)}_{b_1,\cdots,b_k}. \end{aligned}$$

This finishes the proof of Claim 1. □

Let $C^{(k)}_{a_1,\cdots,a_k} = B^{(k)}_{a_1,\cdots,a_k} \cap X$. By the above claim for any $C^{(k)}_{a_1,\cdots,a_k}$ there exists $m \in \mathbb{N}$ such that for each $i = 1, 2$ there are $b_1, \cdots, b_k \in \{1, 2\}$ such that $T^{im} C^{(k)}_{a_1,\cdots,a_k} = C^{(k)}_{b_1,\cdots,b_k}$. Since $\operatorname{diam}(C^{(k)}_{b_1,\cdots,b_k}) \geq 1$ once all $b_1, \cdots, b_k$ belong to $\{1, 2\}$, we see that $m \in N_T(C^{(k)}_{a_1,\cdots,a_k}, \frac{1}{2}) \cap N_{T^2}(C^{(k)}_{a_1,\cdots,a_k}, \frac{1}{2})$. Note that (X, T) is a minimal t.d.s. which is an almost one-to-one extension of the adding machine, by [**22**, Lemma 6.3], for any opene set U there are k and $a_1, \cdots, a_k$ such that $C^{(k)}_{a_1,\cdots,a_k} \subset U$, which implies $N_T(U, \frac{1}{2}) \cap N_{T^2}(U, \frac{1}{2}) \neq \emptyset$. Now applying Lemma 3.3 one has that (X, T) is a $(1, 2)$-sensitive system.

Step 2. Next we show that (X, T) is not $(1, 3)$-sensitive and hence not $(1, 2, 3)$-sensitive.

Denote by $\#(\bullet)$ the cardinality of a set $\bullet$, and for each $m \in \mathbb{N}$ let

$$A^m = \{B^{(4m+2)}_{b_1,\cdots,b_{4m+2}} : b_i \in \{0, 1, 2, 3\} \text{ and } \#(\{i : b_i \in \{0, 3\}\}) \leq m\}.$$

Claim 2. There exists no $k \in \mathbb{N}$ such that $\{T^k B^{(4m+2)}_{0,\cdots,0}, T^{3k} B^{(4m+2)}_{0,\cdots,0}\} \subset A^m$.

Proof of Claim 2. Suppose that there exists $k \in \mathbb{N}$ such that $T^{jk} B^{(4m+2)}_{0,\cdots,0} = B^{(4m+2)}_{a^j_1,\cdots,a^j_{4m+2}} \in A^m$ for some indexes $a^j_i \in \{0, 1, 2, 3\}$ and both $j \in \{1, 3\}$. Since $a^j_i \in \{0, 1, 2, 3\}$ and $\#(\{i : a^j_i \in \{0, 3\}\}) \leq m$ for both $j \in \{1, 3\}$, there exists $l \in \{1, \cdots, 2m, 2m + 1\}$ such that $a^j_{2l-1}, a^j_{2l} \in \{1, 2\}$ for both $j \in \{1, 3\}$. Note that for each $j \in \{1, 3\}$, from the above construction one has $jk = \sum_{i=1}^{4m+2} 4^{i-1} a^j_i \pmod{4^{4m+2}}$ and hence $jk = \sum_{i=1}^{2l} 4^{i-1} a^j_i \pmod{4^{2l}}$. Then there exists $d \in \{0, 1, 2\}$ such that $3(a^1_{2l-1} + 4a^1_{2l}) + d \equiv a^3_{2l-1} + 4a^3_{2l} \pmod{4^2}$. However:

(1) if $a^1_{2l-1} = 1, a^1_{2l} = 1$, then $3(a^1_{2l-1} + 4a^1_{2l}) = 15$;
(2) if $a^1_{2l-1} = 2, a^1_{2l} = 1$, then $3(a^1_{2l-1} + 4a^1_{2l}) = 18$;
(3) if $a^1_{2l-1} = 1, a^1_{2l} = 2$, then $3(a^1_{2l-1} + 4a^1_{2l}) = 27$;
(4) if $a^1_{2l-1} = 2, a^1_{2l} = 2$, then $3(a^1_{2l-1} + 4a^1_{2l}) = 30$.

For all of the above cases, there are no $d \in \{0, 1, 2\}$ and $a^3_{2l-1}, a^3_{2l} \in \{1, 2\}$ which satisfy the required condition. This finishes the proof of Claim 2. □

Observe from the construction of A^m that $\text{diam}(B^{(4m+2)}_{b_1,\cdots,b_{4m+2}}) \leq \frac{1}{2^m}$ for any $B^{(4m+2)}_{b_1,\cdots,b_{4m+2}} \notin A^m$, thus if we put $V_m = B^{(4m+2)}_{0,\cdots,0} \cap X$ then

$$N_T(V_m, \frac{1}{2^m}) \cap N_{T^3}(V_m, \frac{1}{2^m}) = \emptyset.$$

This shows, by the arbitrariness of $m \in \mathbb{N}$, that (X, T) is not $(1, 3)$-sensitive. □

Based on Example 3.7 and Example 3.8, it is natural to expect a construction of a minimal weakly mixing system which is $(1, 2)$-sensitive but not $(1, 2, 3)$-sensitive. While, the following Corollary 3.10 shows that it cannot be done. It also explains why we build Example 3.8 over an odometer, blowing up some fibers.

Lemma 3.9. *Let (X, T) be a minimal t.d.s. If the extension $\phi : (X, T) \to (X_{eq}, T_{eq})$ is not almost one-to-one, then (X, T) is strongly multi-sensitive.*

Proof. Applying [**13**, Theorem 3.1] and [**12**, Theorem 4.2] to the assumptions, there exists $\delta > 0$ such that, for any opene set U, the set $N_T(U, \delta)$ is thickly syndetic and hence $N_{T^i}(U, \delta)$ is also thickly syndetic for each $i \in \mathbb{N}$. Note that the intersection of finitely many thickly syndetic sets is also thickly syndetic, from which we obtain easily that (X, T) is strongly multi-sensitive. □

Since it is trivial the maximal equicontinuous factor (X_{eq}, T_{eq}) of any weakly mixing t.d.s. (X,T), i.e., X_{eq} is a singleton, as a direct corollary of Lemma 3.9 one has:

COROLLARY 3.10. *Any nontrivial minimal weakly mixing system is strongly multi-sensitive.*

4. Weak mixing and Δ-transitivity

In this section we prove that, for a vector $\mathbf{a} = (a_1, \cdots, a_r) \in \mathbb{N}^r$ with $a_1 < \cdots < a_r$ and $r \geq 2$, Δ-$\mathbf{a}$ transitivity implies weak mixing if and only if the vector $\mathbf{a}$ has dependent increments, which follows from Theorem 4.2 and Theorem 4.4.

The concepts of Δ-transitivity and vectors having dependent increments are presented as follows. According to our knowledge, Δ-transitivity was introduced first by Moothathu in [**19**].

DEFINITION 4.1. Let $\mathbf{a} = (a_1, a_2, \cdots, a_r)$ be a vector in $\mathbb{N}^r$ with $r \in \mathbb{N}$ and $a_1 < a_2 < \cdots < a_r$, and put $a_0 = 0$. We say that

1. a t.d.s. (X,T) is Δ-$\mathbf{a}$-*transitive* if there is a point $x \in X$ such that $(x, x, \cdots, x)$ is a transitive point of the system $(X^r, T^{(\mathbf{a})})$.
2. the vector $\mathbf{a}$ has *dependent increments* if $r \geq 2$ and there exists $0 \leq i_1 < i_2 \leq i_3 < i_4 \leq r$ such that $a_{i_4} - a_{i_3} = a_{i_2} - a_{i_1}$.

The easier direction is the following Theorem 4.2.

THEOREM 4.2. *Suppose that (X,T) is a Δ-$\mathbf{a}$-transitive t.d.s., where $\mathbf{a} = (a_1, a_2, \cdots, a_r) \in \mathbb{N}^r, r \in \mathbb{N}$ has dependent increments. Then (X,T) is weakly mixing.*

PROOF. Recall that by convention $a_0 = 0$. As (X,T) is Δ-$\mathbf{a}$-transitive, by [**8**, Proposition 5.3], for any opene sets $U_0, U_1, \cdots, U_r$, there exists $n \in \mathbb{N}$ with $\bigcap_{i=0}^{r} T^{-na_i} U_i \neq \emptyset$. Since $\mathbf{a}$ has dependent increments, there exists $0 \leq i_1 < i_2 \leq i_3 < i_4 \leq r$ such that $a_{i_4} - a_{i_3} = a_{i_2} - a_{i_1}$ and denote $C = a_{i_2} - a_{i_1}$. Then $U_{i_1} \cap T^{-nC} U_{i_2} \neq \emptyset$ and $U_{i_3} \cap T^{-nC} U_{i_4} \neq \emptyset$, thus (X,T) is weakly mixing. □

In order to prove the harder direction, i.e., Theorem 4.4, we will need the following technical Lemma 4.3, whose proof relies on the ideas of [**16**, Lemma 7 and Theorem 8].

We say that a finite set $S \subset \mathbb{N}$ is *q-dispersed*, where $q \geq 2$, if for every $a, b \in S \cup \{0\}$ we have either $a = b$ or $|a - b| \geq q$.

LEMMA 4.3. *Let $\mathbf{a} = (a_1, a_2, \cdots, a_r)$ be a vector in $\mathbb{N}^r$ with $r \in \mathbb{N}$ and $a_1 < a_2 < \cdots < a_r$, which does not have dependent increments. Assume that $A \subset \mathbb{N}$ is an M-dispersed finite set where $M \geq 2$. Then there exists an M-dispersed finite set B containing A such that, $\min(B \setminus A) \geq M + k$ where $k = \max\ A + 1$, and for any $(r+1)$-tuple of words $u_0, u_1, \cdots, u_r$ from $\mathcal{L}_k(\Sigma_B)$, there is $n \in \mathbb{N}$ such that $\bigcap_{i=0}^{r} \sigma_B^{-na_i}([u_i]_B) \neq \emptyset$.*

PROOF. Let $m = \#(\mathcal{L}_k(\Sigma_A))^{r+1}$, i.e., the cardinality of the set of all possible $(r+1)$-tuples of words from $\mathcal{L}_k(\Sigma_A)$. We enumerate all members of this set as a list $U^1, \cdots, U^m$. Hence, each U^j is an ordered list of $(r+1)$-tuples of words from $\mathcal{L}_k(\Sigma_A)$, say $U^j = (u_{j,0}, u_{j,1}, \cdots, u_{j,r})$. Choose $l_1, \cdots, l_m \in \mathbb{N}$ such that

$$l_1 \geq 2k + M - 1 \text{ and } l_{j+1} \geq (a_r + 1) l_j \text{ for all } j = 2, \cdots, m-1.$$

For each $1 \le i \le r$ and $1 \le j \le m$, let $b_i = a_i - a_{i-1}$ (recall $a_0 = 0$) and denote

$$t^j = u_{j,0}0^{b_1 l_j - k}u_{j,1}0^{b_2 l_j - k}u_{j,2}\cdots 0^{b_r l_j - k}u_{j,r}.$$

We put $B = \bigcup_{j=1}^{m} Sp(t^j)$. Recall that $k = \max\ A + 1$. Thus if $n \in A$ then $n < k$, and so $u = 10^{n-1}10^{k-n-1}$ is well defined and clearly $u \in \mathcal{L}_k(\Sigma_A)$, which implies $A \subset B$, because $U^j = (u, 0^k, \cdots, 0^k)$ for some j. Let $E = \{a_j - a_i : 0 \le i < j \le r\}$. Then every element in E has a unique representation, since the vector $\mathbf{a}$ does not have dependent increments. For each j, by the definition of t^j we see that

$$Sp(t^j) \setminus A \subset \bigcup_{i\in E} C_i^j, \text{ where } C_i^j = [il_j - k + 1, il_j + k - 1] \cap \mathbb{N} \text{ for each } i \in E.$$

Thus $\min(B \setminus A) \ge l_1 - k + 1 \ge M + k$, and then $\min\ B = \min\ A \ge M$. We will show that, if $r \in B \setminus A$, then there are unique indexes $j(r), i(r)$ such that $r \in Sp(t^{j(r)})$ and $r \in C_{i(r)}^{j(r)}$. First observe that for any $j' > j$ and $b \in E$ we have

$$\begin{aligned} bl_{j'} - k &\ge bl_{j+1} - k \ge b(a_r + 1)l_j - k \ge a_r l_j + k \\ &\ge (b_1 + \ldots + b_r + 1)l_j + k = |t_j| \end{aligned}$$

and therefore $\min(Sp(t^{j'}) \setminus A) > \max(Sp(t^j) \setminus A)$. This shows that $j(r)$ is uniquely determined. Similarly, if $i' > i$ then

$$i'l_j - k + 1 \ge il_j + l_1 - k + 1 \ge il_j + k + M > il_j + k - 1$$

which shows $\min(C_{i'}^j) > \max(C_{i'}^j)$ and therefore $i(r)$ is also unique.

Now for any $(r+1)$-tuple of words $u_0, u_1, \cdots, u_r$ from $\mathcal{L}_k(\Sigma_B)$, we are going to find $n \in \mathbb{N}$ such that $\bigcap_{i=0}^{r} \sigma_B^{-na_i}([u_i]_B) \neq \emptyset$. In fact, $\mathcal{L}_k(\Sigma_B) = \mathcal{L}_k(\Sigma_A)$, since $\min(B \setminus A) \ge M + k$ and $A \subset B$. Thus there exists j such that $U^j = (u_{j,0}, u_{j,1}, \cdots, u_{j,r}) = (u_0, u_1, \cdots, u_r)$, and then by the construction $t^j = u_0 0^{b_1 l_j - k}u_1 0^{b_2 l_j - k}u_2 \cdots 0^{b_r l_j - k}u_r \in \mathcal{L}(\Sigma_B)$, which implies

$$t^j 0^\infty \in \bigcap_{i=0}^{r} \sigma_B^{-a_i l_j}([u_j]_B) \neq \emptyset.$$

Finally we show that B is M-dispersed. It suffices to prove $|q - p| \ge M$ for all $q, p \in B$ with $q \neq p$. The case that both p and q belong to A is clear, since A is M-dispersed. The case that one of p, q belongs to A and the other belongs to $B \setminus A$ follows from the fact that $\min(B \setminus A) \ge M + k > M + \max\ A$. Now we consider the remaining case that both p and q belong to $B \setminus A$. There are three sub-cases according to $j(p), i(p)$ and $j(q), i(q)$.

Case I: $j(p) \neq j(q)$. Without loss of generality we assume $j(q) > j(p)$, and then

$$\begin{aligned} q &\ge l_{j(q)} - k + 1 \ge (a_r + 1)l_{j(p)} - k + 1 \\ &\ge a_r l_{j(p)} + l_1 - k + 1 \ge a_r l_{j(p)} + k + M \ge p + M. \end{aligned}$$

Case II: $j(p) = j(q) = j$, but $i(p) \neq i(q)$. Similarly we may assume $i(q) > i(p)$, thus

$$\begin{aligned} q &\ge i(q)l_j - k + 1 \ge (i(p) + 1)l_j - k + 1 \\ &\ge i(p)l_j - k + 1 + l_1 \ge i(p)l_j + k + M \ge p + M. \end{aligned}$$

Case III: $j(p) = j(q)$ $(= j)$ and $i(p) = i(q)$ $(= i)$. For each $r \in \{p, q\}$, we define

$$s(r) = \min\{s : (t^j)_s = (t^j)_{s+r} = 1\}.$$

Clearly, either $s(p) \neq s(q)$, or $s(p)+p \neq s(q)+q$. Let $c = |(s(p)+p)-(s(q)+q)|$ and $d = |s(p) - s(q)|$. Since $\mathbf{a}$ does not have dependent increments, by the construction of t^j we have $c, d \in A \cup \{0\}$. Thus if $c = d \neq 0$, then $|q - p| = 2d \geq 2M$; and if $c \neq d$, then $|q - p| \geq |c - d| \geq M$. This finishes the proof. □

Now we are ready to prove the following result.

THEOREM 4.4. *Let $\mathbf{a} = (a_1, a_2, \cdots, a_r)$ be a vector in $\mathbb{N}^r$ with $r \in \mathbb{N}$ and $a_1 < a_2 < \cdots < a_r$, which does not have dependent increments. Then there exists $P \subset \mathbb{N}$ such that the spacing shift (Σ_P, σ_P) is Δ-$\mathbf{a}$-transitive but not weakly mixing.*

PROOF. Fix any integer $M \geq 2$ and let $P_0 = \{M\}$. Define inductively a sequence $P_n \subset \mathbb{N}, n \in \mathbb{N}$, where each P_n is provided by Lemma 4.3 for $A = P_{n-1}$ and M. Define $P = \bigcup_{n=0}^{\infty} P_n$.

By the construction $P_0 \subset P_1 \subset P_2 \cdots$, we have $|p - q| \geq M$ for every distinct $p, q \in P$. In particular, $P \subset \mathbb{N}$ is not thick, and so the spacing shift (Σ_P, σ_P) is not weakly mixing.

Now consider any opene sets $U_0, U_1, \cdots, U_r \subset \Sigma_P$. Clearly there exists $k = \max P_l + 1$ (for some $l \in \mathbb{N}$) and words $u_0, u_1, \cdots u_r \in \mathcal{L}_k(\Sigma_P)$ such that $[u_i]_P \subset U_i$ for all $i = 0, 1, \cdots, r$. Note that $\min\,(P_{l+1} \setminus P_l) \geq M + k$ by Lemma 4.3, and hence $u_0, u_1, \cdots u_r \in \mathcal{L}_k(\Sigma_{P_{l+1}})$, applying again Lemma 4.3 we may choose $n \in \mathbb{N}$ such that

$$\bigcap_{i=0}^{r} \sigma_P^{-a_i n}(U_i) \supset \bigcap_{i=0}^{r} \sigma_P^{-a_i n}([u_i]_P) \supset \bigcap_{i=0}^{r} \sigma_{P_{l+1}}^{-a_i n}([u_i]_{P_{l+1}}) \neq \emptyset.$$

Thus the spacing shift (Σ_P, σ_P) is Δ-$\mathbf{a}$-transitive. □

5. Multi-ergodicity

By a *measurable dynamical system* (*m.d.s.* for short) we mean a quadruple $(X, \mathcal{B}, \mu, T)$, where $(X, \mathcal{B}, \mu)$ is a Lebesgue space and $T : (X, \mathcal{B}, \mu) \to (X, \mathcal{B}, \mu)$ is a measure-preserving invertible transformation. Recall that a m.d.s. $(X, \mathcal{B}, \mu, T)$ is *weakly mixing* if the product m.d.s. $(X, \mathcal{B}, \mu, T) \times (X, \mathcal{B}, \mu, T)$ is ergodic.

Note that ergodic theory and topological dynamics exhibit a remarkable parallelism. It is natural to introduce for measurable dynamical systems a notion similar to multi-transitivity, which we call it *multi-ergodicity* as follows.

DEFINITION 5.1. Let $\mathbf{a} = (a_1, a_2, \cdots, a_r) \in \mathbb{N}^r, r \in \mathbb{N}$. We say that a m.d.s. $(X, \mathcal{B}, \mu, T)$ is

(1) $\mathbf{a}$-*ergodic* if the product m.d.s. $\prod_{i=1}^{r} (X, \mathcal{B}, \mu, T^{a_i})$ is ergodic.
(2) *strongly multi-ergodic* if it is $\mathbf{a}$-ergodic for any vector $\mathbf{a} \in \mathbb{N}^n, n \in \mathbb{N}$.

Though in [**16**] Kwietniak and Oprocha showed that in general there is no connection between multi-transitivity and weak mixing, Moothathu proved that for a minimal system multi-transitivity is equivalent to weak mixing [**19**]. And so we may ask how about the relationship between multi-ergodicity and weak mixing for measurable dynamical systems.

Before proceeding, we need the following result from topological dynamics, which was proved firstly for the vector $(1,2)$ by Moothathu [**19**]. Note that though the following result holds for t.d.s. with a general compact Hausdorff space, we will only present its proof in the compact metrizable setting.

LEMMA 5.2. *Let (X,T) be a minimal t.d.s. and $a,b \in \mathbb{N}$. If the product t.d.s. $(X \times X, T^a \times T^b)$ is transitive, then (X_{eq}, T_{eq}) is a trivial system, i.e., X_{eq} is a singleton.*

PROOF. Since (X_{eq}, T_{eq}) is a minimal equicontinuous t.d.s., there is an equivalent metric on X_{eq}, denoted still be d without any confusion, under which T_{eq} is an isometry [**2**]. For simplicity of notation denote $S = T_{eq}$. Since $(X_{eq} \times X_{eq}, S^a \times S^b)$ is transitive, there exists $(y_1, y_2) \in X_{eq} \times X_{eq}$ with a dense orbit for $S^a \times S^b$ in $X_{eq} \times X_{eq}$.

Now fix each $x \in X_{eq}$ and any $\varepsilon > 0$. There are $n, m_1, m_2 > 0$ such that

$$d(S^{an}y_1, y_1) < \varepsilon, \quad d(S^{bn}y_2, Sy_2) < \varepsilon, \quad d(S^{m_1}x, y_1) < \varepsilon, \quad d(S^{m_2}x, y_2) < \varepsilon.$$

Then

$$\begin{aligned} d(x, S^{an}x) &= d(S^{m_1}x, S^{m_1}S^{an}x) \\ &\leq d(S^{m_1}x, y_1) + d(y_1, S^{an}y_1) + d(S^{an}y_1, S^{an}S^{m_1}x) < 3\varepsilon \end{aligned}$$

and

$$\begin{aligned} d(Sx, S^{bn}x) &= d(S^{m_2+1}x, S^{m_2}S^{bn}x) \\ &\leq d(S^{m_2+1}x, Sy_2) + d(Sy_2, S^{bn}y_2) + d(S^{bn}y_2, S^{bn}S^{m_2}x) < 3\varepsilon. \end{aligned}$$

We obtain $d(S^{abn}x, x) < 3b\varepsilon$ and $d(S^{abn}x, S^a x) < 3a\varepsilon$, which gives $d(x, S^a x) < 3(a+b)\varepsilon$. Since x and ε are arbitrary, the transformation S^a is the identity map. But (X_{eq}, S^a) is a transitive t.d.s., one has that X_{eq} must be a singleton. □

Let $(X, \mathcal{B}, \mu, T)$ be a m.d.s. It is well known that there exists a T-invariant σ-algebra $\mathcal{K}_\mu \subset \mathcal{B}$ such that $L^2(X, \mathcal{K}_\mu, \mu)$ is exactly the closure in $L^2(X, \mathcal{B}, \mu)$ of linear span of

$$\{f \in L^2(X, \mathcal{B}, \mu) : \exists \lambda \in \mathbb{C} \text{ s.t. } f \circ T = \lambda f\}$$

(see for example [**23**, Theorem 7.1]). Furthermore, $(X, \mathcal{B}, \mu, T)$ has discrete spectrum if and only if $\mathcal{K}_\mu = \mathcal{B}$; and is weakly mixing if and only if $\mathcal{K}_\mu$ is trivial, i.e., $\mathcal{K}_\mu = \{\emptyset, X\}$. In particular, $(X, \mathcal{K}_\mu, \mu, T)$ has discrete spectrum. We are reluctant to present definition of discrete spectrum for a m.d.s., and we refer the reader to [**21**] for more details.

PROPOSITION 5.3. *Let $(X, \mathcal{B}, \mu, T)$ be a m.d.s. and $a, b \in \mathbb{N}$. If $(X \times X, \mathcal{B} \times \mathcal{B}, \mu \times \mu, T^a \times T^b)$ is ergodic, then $(X, \mathcal{B}, \mu, T)$ is weakly mixing.*

PROOF. Since $(X, \mathcal{B}, \mu, T^a)$ is ergodic, $(X, \mathcal{B}, \mu, T)$ is also ergodic, and hence the above-mentioned m.d.s. $(X, \mathcal{K}_\mu, \mu, T)$ is ergodic and has discrete spectrum. By the well-known Halmos-von Neumann Representation Theorem (see for example [**10**] or [**21**, Theorem 3.6]), we may view $(X, \mathcal{K}_\mu, \mu, T)$ as $(G, \mathcal{C}, \nu, S)$, where (G, S) is a minimal rotation over a compact abelian metric group with $\mathcal{C}$ and ν its Borel σ-algebra and its normalized Haar measure, respectively. In particular, (G, S) is a minimal equicontinuous t.d.s.

As $(X \times X, \mathcal{B} \times \mathcal{B}, \mu \times \mu, T^a \times T^b)$ is ergodic, one has similarly that $(G \times G, \mathcal{C} \times \mathcal{C}, \nu \times \nu, S^a \times S^b)$ is also ergodic, and then the t.d.s. $(G \times G, S^a \times S^b)$ is transitive. Therefore by Lemma 5.2 we have that G is a singleton (note that it is itself the maximal equicontinuous factor of (G, S)). Equivalently, $\mathcal{K}_\mu$ is trivial, and so $(X, \mathcal{B}, \mu, T)$ is weakly mixing. □

Now we are ready to present the main result of this section that multi-ergodicity and weak mixing are equivalent for a m.d.s.

THEOREM 5.4. *Let $(X, \mathcal{B}, \mu, T)$ be a m.d.s. The following conditions are equivalent:*

(1) $(X, \mathcal{B}, \mu, T)$ *is weakly mixing.*
(2) $(X \times X, \mathcal{B} \times \mathcal{B}, \mu \times \mu, T^a \times T^b)$ *is ergodic for some* $a, b \in \mathbb{N}$*.*
(3) $(X, \mathcal{B}, \mu, T)$ *is strongly multi-ergodicity.*

PROOF. The implication (2) ⇒ (1) follows from Proposition 5.3. The implication (3) ⇒ (2) is trivial. Now let us prove (1) ⇒ (3). We assume that $(X, \mathcal{B}, \mu, T)$ is weakly mixing, and let $\mathbf{a} = (a_1, a_2, \cdots, a_r) \in \mathbb{N}^r, r \in \mathbb{N}$. We need to show that $\prod_{i=1}^{r} (X, \mathcal{B}, \mu, T^{a_i})$ is ergodic.

We prove it by induction for $r \in \mathbb{N}$. Note that by [**9**, Proposition 4.7], the m.d.s. $(X, \mathcal{B}, \mu, T^k)$ is weakly mixing and hence ergodic for each $k \in \mathbb{N}$, which finishes the proof for the case of $r = 1$. Now assume that $\prod_{i=1}^{r} (X, \mathcal{B}, \mu, T^{a_i})$ is ergodic, and let $a_{r+1} \in \mathbb{N}$. Since $(X, \mathcal{B}, \mu, T^{a_{r+1}})$ is weakly mixing, applying [**9**, Proposition 4.5] the product system $\prod_{i=1}^{r+1} (X, \mathcal{B}, \mu, T^{a_i})$ is also ergodic, which ends the proof. □

Acknowledgments

This work was partially done during P. Oprocha's visit to the School of Mathematical Sciences, Fudan University at the beginning of 2018. He gratefully acknowledges the hospitality of Fudan University.

References

[1] Ethan Akin, *Recurrence in topological dynamics*, The University Series in Mathematics, Plenum Press, New York, 1997. Furstenberg families and Ellis actions. MR1467479
[2] Ethan Akin, Joseph Auslander, and Kenneth Berg, *When is a transitive map chaotic?*, Convergence in ergodic theory and probability (Columbus, OH, 1993), Ohio State Univ. Math. Res. Inst. Publ., vol. 5, de Gruyter, Berlin, 1996, pp. 25–40. MR1412595
[3] Joseph Auslander, *Minimal flows and their extensions*, North-Holland Mathematics Studies, vol. 153, North-Holland Publishing Co., Amsterdam, 1988. Notas de Matemática [Mathematical Notes], 122. MR956049
[4] Joseph Auslander and James A. Yorke, *Interval maps, factors of maps, and chaos*, Tôhoku Math. J. (2) **32** (1980), no. 2, 177–188, DOI 10.2748/tmj/1178229634. MR580273
[5] John Banks, *Regular periodic decompositions for topologically transitive maps*, Ergodic Theory Dynam. Systems **17** (1997), no. 3, 505–529, DOI 10.1017/S0143385797069885. MR1452178
[6] John Banks, Thi T. D. Nguyen, Piotr Oprocha, Brett Stanley, and Belinda Trotta, *Dynamics of spacing shifts*, Discrete Contin. Dyn. Syst. **33** (2013), no. 9, 4207–4232, DOI 10.3934/dcds.2013.33.4207. MR3038059
[7] ZhiJing Chen, Jian Li, and Jie Lü, *On multi-transitivity with respect to a vector*, Sci. China Math. **57** (2014), no. 8, 1639–1648, DOI 10.1007/s11425-014-4797-z. MR3229228
[8] Zhijing Chen, Jian Li, and Jie Lü, *Point transitivity, Δ-transitivity and multi-minimality*, Ergodic Theory Dynam. Systems **35** (2015), no. 5, 1423–1442, DOI 10.1017/etds.2013.106. MR3365729

[9] H. Furstenberg, *Recurrence in ergodic theory and combinatorial number theory*, Princeton University Press, Princeton, N.J., 1981. M. B. Porter Lectures. MR603625

[10] Paul R. Halmos and John von Neumann, *Operator methods in classical mechanics. II*, Ann. of Math. (2) **43** (1942), 332–350, DOI 10.2307/1968872. MR0006617

[11] Wen Huang, Danylo Khilko, Sergiĭ Kolyada, Alfred Peris, and Guohua Zhang, *Finite intersection property and dynamical compactness*, J. Dynam. Differential Equations **30** (2018), no. 3, 1221–1245, DOI 10.1007/s10884-017-9600-8. MR3842147

[12] Wen Huang, Danylo Khilko, Sergiĭ Kolyada, and Guohua Zhang, *Dynamical compactness and sensitivity*, J. Differential Equations **260** (2016), no. 9, 6800–6827, DOI 10.1016/j.jde.2016.01.011. MR3461085

[13] Wen Huang, Sergiĭ Kolyada, and Guohua Zhang, *Analogues of Auslander-Yorke theorems for multi-sensitivity*, Ergodic Theory Dynam. Systems **38** (2018), no. 2, 651–665, DOI 10.1017/etds.2016.48. MR3774836

[14] Lixin Jiao, Lidong Wang, Fengquan Li, and Heng Liu, *On multi-sensitivity with respect to a vector*, Modern Phys. Lett. B **32** (2018), no. 15, 1850166, 13, DOI 10.1142/S021798491850166X. MR3805946

[15] Petr Kůrka, *Topological and symbolic dynamics*, Cours Spécialisés [Specialized Courses], vol. 11, Société Mathématique de France, Paris, 2003. MR2041676

[16] Dominik Kwietniak and Piotr Oprocha, *On weak mixing, minimality and weak disjointness of all iterates*, Ergodic Theory Dynam. Systems **32** (2012), no. 5, 1661–1672, DOI 10.1017/S0143385711000599. MR2974213

[17] Kenneth Lau and Alan Zame, *On weak mixing of cascades*, Math. Systems Theory **6** (1972/73), 307–311, DOI 10.1007/BF01740722. MR0321058

[18] T. K. Subrahmonian Moothathu, *Stronger forms of sensitivity for dynamical systems*, Nonlinearity **20** (2007), no. 9, 2115–2126, DOI 10.1088/0951-7715/20/9/006. MR2351026

[19] T. K. Subrahmonian Moothathu, *Diagonal points having dense orbit*, Colloq. Math. **120** (2010), no. 1, 127–138, DOI 10.4064/cm120-1-9. MR2652611

[20] David Ruelle, *Dynamical systems with turbulent behavior*, Mathematical problems in theoretical physics (Proc. Internat. Conf., Univ. Rome, Rome, 1977), Lecture Notes in Phys., vol. 80, Springer, Berlin-New York, 1978, pp. 341–360. MR518445

[21] Peter Walters, *An introduction to ergodic theory*, Graduate Texts in Mathematics, vol. 79, Springer-Verlag, New York-Berlin, 1982. MR648108

[22] Tao Yu, *Measurable sensitivity via Furstenberg families*, Discrete Contin. Dyn. Syst. **37** (2017), no. 8, 4543–4563, DOI 10.3934/dcds.2017194. MR3642276

[23] Robert J. Zimmer, *Ergodic actions with generalized discrete spectrum*, Illinois J. Math. **20** (1976), no. 4, 555–588. MR0414832

AGH University of Science and Technology, Faculty of Applied Mathematics, al. Mickiewicza 30, 30-059 Kraków, Poland; and National Supercomputing Centre IT4Innovations, Division of the University of Ostrava, Institute for Research and Applications of Fuzzy Modeling, 30. dubna 22, 70103 Ostrava, Czech Republic

Email address: oprocha@agh.edu.pl

Shanghai Center for Mathematical Sciences, Fudan University, Shanghai 200438, China

Email address: ytnuo@mail.ustc.edu.cn

School of Mathematical Sciences and Shanghai Center for Mathematical Sciences, Fudan University, Shanghai 200433, China

Email address: chiaths.zhang@gmail.com

Contemporary Mathematics
Volume **744**, 2020
https://doi.org/10.1090/conm/744/14927

Convergence of zeta functions for amenable group extensions of shifts

Richard Sharp

Dedicated to the memory of Sergiy Kolyada

Abstract. The aim of this paper is show how zeta functions for amenable group extensions of subshifts of finite type may be approximated by rescaled zeta functions for a sequence of subshifts of finite type whose states are determined by a Følner exhaustion of the group. In particular, this recovers a result of Guido, Isola and Lapidus for graphs, and, by using weighted zeta functions, extends it to metric graphs.

1. Introduction

A classical object attached to a finite graph X is its *Ihara zeta function*

$$\zeta_X(z) = \prod_\gamma (1 - z^{|\gamma|})^{-1}, \tag{1.1}$$

where γ runs over the prime closed geodesics of X and $|\gamma|$ denotes the length of the geodesic. (Here, a closed geodesic is a closed path in X, with no backtracking or tail, modulo cyclic permutation. It is prime if it traverses its image exactly once. Its length is the number of edges forming the path.) The product converges for $|z|$ sufficiently small and extends to a rational function, which may be expressed as a determinant in various ways. See [**21**] for an account of this theory.

Recently, there has also been considerable interest in the analogous theory for infinite graphs, particularly those which occur as covers of finite graphs [**3–5**, **9–11**, **15**]. It is clear that, in this case, a naïve definition analogous to (1.1) does not work, since, for example, if one considers an infinite regular cover of a finite graph then any prime closed geodesic will have infinitely many translates with the same length. It is these covers that we will be concerned with and a natural way to proceed is to define a zeta function relative to the cover, i.e. to take a product over prime closed geodesics in the cover modulo the action of the covering group. An alternative approach, discussed in [**9**], is to define the zeta function for the cover as a limit of normalised zeta functions for finite graphs. For example, if the covering group is residually finite, then one might consider the zeta functions of finite subcovers.

2010 *Mathematics Subject Classification.* Primary 05C63 37C25 37C30.

Key words and phrases. Zeta functions, subshifts of finite types, skew products, amenable groups, infinite graphs.

The author was supported by the United Kingdom Engineering and Physical Sciences Research Council grant number EP/P028373/1.

On the other hand, if the covering group is amenable, then one might consider finite graphs associated to an exhausting Følner sequence. This second situation is studied in [**11**], where it is shown that, in an appropriate disk, the normalised zeta functions for a Følner sequence converge to the zeta function for the infinite graph relative to the covering. The aim of this note is to extend this result to zeta functions for certain dynamical systems; precisely, amenable group extensions of subshifts of finite type. We are also able to consider weighted zeta functions and hence, as an application, zeta functions for metric graphs.

We will now describe our set-up in more detail. Let $\mathcal{S}$ be a finite set of states with the discrete topology and let A be a $\#\mathcal{S} \times \#\mathcal{S}$ zero-one matrix indexed by $\mathcal{S}$. We then define

$$\Sigma = \left\{(x_n)_{n=0}^{\infty} \in \mathcal{S}^{\mathbb{Z}^+} : A(x_n, x_{n+1}) = 1 \ \forall n \geq 0\right\}$$

with the subspace topology, where $\mathcal{S}^{\mathbb{Z}^+}$ has the product topology, making Σ totally disconnected. If we let $\sigma : \Sigma \to \Sigma$ denote the left shift map then the pair (Σ, σ) is called a subshift of finite type and we call A the transition matrix. (Strictly speaking, (Σ, σ) is a *one-sided* subshift of finite type and there is a corresponding two-sided subshift of finite type defined on the corresponding subset of $\mathcal{S}^{\mathbb{Z}}$. However, as we are only interested in periodic orbits, it is no restriction to limit ourselves to the one-sided case.) We will also assume that $\sigma : \Sigma \to \Sigma$ is topologically transitive, which is equivalent to the requirement that, for each $(i,j) \in \mathcal{S} \times \mathcal{S}$, there is an $n = n(i,j) \geq 1$ such that $A^n(i,j) > 0$.

We say that a periodic orbit $\tau = \{x, \sigma x, \ldots, \sigma^{n-1}x\}$, with $\sigma^n x = x$, is *prime* if $\sigma^m x \neq x$ for $1 \leq m < n$ and write $|\tau| = n$. Let $\mathcal{P}(\sigma)$ denote the set of prime periodic orbits of σ. We define the *Artin–Mazur zeta function* for σ to be the function

$$\zeta_\sigma(z) := \prod_{\tau \in \mathcal{P}(\sigma)} (1 - z^{|\tau|})^{-1} = \exp \sum_{n=1}^{\infty} \frac{\#\{x \in \Sigma : \sigma^n x = x\}}{n} z^n. \tag{1.2}$$

(The equality of the two expressions is a standard result.) This function converges for $|z| < \exp(-h(\sigma))$, where $h(\sigma)$ is the topological entropy of σ, and has a rational extension to the whole complex plane given by

$$\zeta_\sigma(z) = \frac{1}{\det(I - zA)}. \tag{1.3}$$

We now consider extensions by countable amenable groups. (In fact, we will only consider finitely generated groups.) There are numerous equivalent definitions of amenability. We shall use the convenient characterisation, due to Følner [**8**], that a countable group G is amenable if for every finite set $F \subset G$ and every $\epsilon > 0$, there exists a finite set $K \subset G$ such that $\#(F \triangle Fg) < \epsilon \#F$, for all $g \in F$.

Let G be a finitely generated amenable group (with the discrete topology) and suppose that we are given a continuous function $\psi : \Sigma \to G$. Since G is discrete, the continuity of ψ implies that ψ is locally constant, i.e. that there exists $N \geq 1$ such that $\psi(x) = \psi(x_0, x_1, \ldots, x_{N-1})$, for $x = (x_i)_{i=0}^{\infty} \in \Sigma$. The following standard recoding allows us to assume the $N = 2$. We can define a new subshift of finite type (Σ', σ') with state set

$$\mathcal{S}' := \{(i_0, i_1, \ldots, i_{N-2}) \in \mathcal{S}^{N-1} : A(i_n, i_{n+1}) = 1, \, n = 0, \ldots, N-3\}$$

and transition matrix A' given by

$$A'((i_0,\dots,i_{N-2}),(j_0,\dots,j_{N-2})) = \begin{cases} 1 & \text{if } j_n = i_{n+1},\ n=0,\dots,N-3, \\ & \text{and } A(i_{N-2},j_{N-2}) = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Then the map $((i_0,\dots,i_{N-2}),(i_1,\dots,i_{N-1}),\dots) \mapsto (i_0,i_1,\dots)$ from Σ' to Σ is a topological conjugacy and, in particular, we may identify the two sets of periodic orbits. Furthermore, the conjugacy takes ψ to a function on Σ' that depends on two co-ordinates. Therefore, in what follows, it is no loss of generality to suppose that $\psi : \Sigma \to G$ depends on only two co-ordinates.

The function ψ defines a skew product extension $T_\psi : \Sigma \times G \to \Sigma \times G$ by

$$T_\psi(x,g) = (\sigma x, g\psi(x)).$$

We will always assume that T_ψ is topologically transitive, which, in particular, implies that $\Psi = \{\psi(i,j) : i,j \in \mathcal{S},\ A(i,j)=1\}$ is a generating set for G, i.e. that every element of G may be written as a product of elements of Ψ and their inverses.

The map $T_\psi : \Sigma \times G \to \Sigma \times G$ is itself a countable state subshift with infinite transition matrix $\mathbb{A}$, indexed by $\mathcal{S} \times G$, defined by

$$\mathbb{A}((i,g),(j,h)) = \begin{cases} 1 & \text{if } A(i,j) = 1 \text{ and } h = g\psi(i,j), \\ 0 & \text{otherwise.} \end{cases}$$

We note that $T_\psi^n(x,g) = (x,g)$ if and only if $\sigma^n x = x$ and

$$\psi^n(x) := \psi(x)\psi(\sigma x)\cdots\psi(\sigma^{n-1}x) = e,$$

the identity in G. In particular, $T_\psi^n(x,g) = (x,g)$ for some $g \in G$ if and only if $T_\psi^n(x,g) = (x,g)$ for all $g \in G$.

As for σ, we write $\mathcal{P}(T_\psi)$ for the set of prime T_ψ-periodic orbits. (Note that if $T_\psi^n(x,g) = (x,g)$ then $\sigma^n x = x$ but that, if G has torsion, the periodic orbit $\{x,\sigma x,\dots,\sigma^{n-1}\}$ is not necessarily prime.) The group G acts transitively on $\mathcal{P}(T_\psi)$ and we write $\mathcal{P}_G(T_\psi) = \mathcal{P}(T_\psi)/G$. For $[\tau] \in \mathcal{P}_G(T_\psi)$, we write $|[\tau]| = |\tau|$, for any $\tau \in [\tau]$; clearly, this is well-defined. The (Artin–Mazur) zeta function for T_ψ is defined in this setting to be

(1.4)
$$\zeta_{T_\psi}(z) = \prod_{[\tau]\in\mathcal{P}_G(T_\psi)} (1 - z^{|[\tau]|})^{-1} = \exp\sum_{n=1}^\infty \frac{\#\{x\in\Sigma : \sigma^n x = x,\ \psi^n(x) = e\}}{n} z^n.$$

An alternative definition for the zeta function of an infinite graph appears in the work of Chinta, Jorgenson and Karlsson [**2**]. This is obtained by restricting to closed paths based at a given vertex. The natural analogue for T_ψ is to set, for $a \in \mathcal{S}$,

$$\zeta_{T_\psi}^{(a)}(z) = \exp\sum_{n=1}^\infty \frac{\#\{x\in\Sigma : \sigma^n x = x,\ x_0 = a,\ \psi^n(x) = e\}}{n} z^n.$$

The radius of convergence of all these functions may be given in terms of the *Gurevič entropy* $h_{\mathrm{Gur}}(T_\psi)$ of T_ψ [**12**]. This is defined by

$$h_{\mathrm{Gur}}(T_\psi) = \limsup_{n\to\infty} \frac{1}{n}\#\{x\in\Sigma : \sigma^n x = x,\ x_0 = a,\ \psi^n(x) = e\},$$

(which is independent of the choice of a by transitivity) and so we see that each $\zeta_{T_\psi}^{(a)}(z)$ has radius of convergence $\exp(-h_{\mathrm{Gur}}(T_\psi))$. Since

$$\#\{x \in \Sigma : \sigma^n x = x,\ x_0 = a,\ \psi^n(x) = e\} \leq \#\{x \in \Sigma : \sigma^n x = x,\ \psi^n(x) = e\}$$
$$\leq \sum_{a \in \mathcal{S}} \#\{x \in \Sigma : \sigma^n x = x,\ x_0 = a,\ \psi^n(x) = e\},$$

$\zeta_{T_\psi}(z)$ has the same radius of convergence.

Remark 1.1. It is interesting to compare the radii of convergence of $\zeta_\sigma(z)$ and $\zeta_{T_\psi}(z)$. As a direct consequence of the definitions of topological entropy and Gurevič entropy, it is clear that $h_{\mathrm{Gur}}(T_\psi) \leq h(\sigma)$, and so the radius of convergence of $\zeta_{T_\psi}(z)$ is at least as large as the radius of convergence of $\zeta_\sigma(z)$. The question of when we have equality reduces to a question about the skew product induced by T_ψ on $\Sigma \times G^{\mathrm{ab}}$, where $G^{\mathrm{ab}} := G/[G,G]$ is the abelianisation of G. More precisely, let $\pi : G \to G^{\mathrm{ab}}$ be the natural projection and define $\varphi : \Sigma \to G^{\mathrm{ab}}$ by $\varphi = \pi \circ \psi$. This induces a skew product $T_\varphi : \Sigma \times G^{\mathrm{ab}} \to \Sigma \times G^{\mathrm{ab}}$. Since G is amenable, $h_{\mathrm{Gur}}(T_\psi) = h_{\mathrm{Gur}}(T_\varphi)$ [**7**]. This means we need to compare $h(\sigma)$ with $h_{\mathrm{Gur}}(T_\varphi)$. If G^{ab} is finite then it is trivial that $h_{\mathrm{Gur}}(T_\varphi) = h(\sigma)$. On the other hand, if G^{ab} is infinite then it is a finite extension of $\mathbb{Z}^d$, for some $d \geq 1$, and we can factor out the finite group to get a function $\varphi_0 : \Sigma \to \mathbb{Z}^d$. This in turn gives a skew product $T_{\varphi_0} : \Sigma \times \mathbb{Z}^d \to \Sigma \times \mathbb{Z}^d$ with $h_{\mathrm{Gur}}(T_{\varphi_0}) = h_{\mathrm{Gur}}(T_\varphi)$. Finally, it follows from the results of [**17**] that $h_{\mathrm{Gur}}(T_{\varphi_0}) = h(\sigma)$ if and only if $\int \varphi_0 \, d\mu_0 = 0$, where μ_0 is the measure of maximal entropy for σ.

We now wish to consider finite approximations to T_ψ, obtained by restricting to a (large) finite subset of G. More precisely, if K is a finite subset of G then we may obtain a subshift of finite type by restricting $\mathbb{A}$ to $\mathcal{S} \times K$, i.e. we consider the finite matrix A_K, indexed by $\mathcal{S} \times K$ and defined by $A_K((i,g),(j,h)) = 1$ if and only if $\mathbb{A}((i,g),(j,h)) = 1$. We write $\sigma_K : \Sigma_K \to \Sigma_K$ for the subshift of finite type with transition matrix A_K and $\zeta_{\sigma_K}(z)$ for the associated zeta function.

We say that a sequence of finite sets $K_n \subset G$, $n \geq 1$, is a Følner exhaustion of G if

(FE1) $\bigcup_{n=1}^\infty K_n = G$;
(FE2) $K_n \subset K_{n+1}$, for all $n \geq 1$; and
(FE3)
$$\lim_{n\to\infty} \frac{\#(K_n \triangle K_n g)}{\# K_n} = 0,$$
for all $g \in G$.

It is easy to see that amenability is equivalent to the existence of a sequence satisfying (FE2) and (FE3). Moreover, it is well-known that the sequence can also be chosen to satisfy (FE1) (see, for example, Theorem 6.2 of [**11**], applied to the action of G on its own Cayley graph, for a proof). Thus G is amenable if and only if it admits a Følner exhaustion.

We have the following convergence result.

Theorem 1.2. *Let $\sigma : \Sigma \to \Sigma$ be a subshift of finite type and let $T_\psi : \Sigma \to G \to \Sigma \times G$ be a topologically transitive skew product extension, where G is a finitely generated amenable group. Suppose that K_n, $n \geq 1$ is a Følner exhaustion of G.*

Then we have

$$\zeta_{T_\psi}(z) = \lim_{n\to\infty} \zeta_{\sigma_{K_n}}(z)^{1/(\#K_n)},$$

uniformly on compact subsets of $\{z \in \mathbb{C} : |z| < (2\|\mathbb{A}\|)^{-1}\}$*, where* $\|\mathbb{A}\|$ *is the operator norm of* $\mathbb{A}$ *acting on* $\ell^2(\mathcal{S} \times G)$*.*

REMARK 1.3. We note that $\|\mathbb{A}\| \le \sqrt{d_r d_c}$, where

$$d_r = \max_{i\in\mathcal{S}} \sum_{j\in\mathcal{S}} A(i,j) \quad \text{and} \quad d_c = \max_{j\in\mathcal{S}} \sum_{i\in\mathcal{S}} A(i,j),$$

i.e. d_r and d_c are, respectively, the maximum row and column sums of A. See Lemma 4.1 for a proof.

I am grateful to Stephen Cantrell for his comments on a draft version of this paper. I am also grateful to the referee for numerous suggestion which have improved the presentation.

2. Graphs

In this section, we show how zeta functions for graphs may be interpreted in the framework of the preceding section and compare our results with those of [**11**]. Let X be a finite connected graph with vertex set $V(X)$ and edge set $E(X)$. (We allow loops and multiple edges.) A closed geodesic is a closed path in X, with no backtracking or tail, modulo cyclic permutation, and a closed geodesic is prime if it traverses its image exactly once. The Ihara zeta function $\zeta_X(z)$ is defined by (1.1) and extends to a rational function via the determinant formula of Bass [**1**] (see also [**14, 21**]):

$$\zeta_X(z) = (1 - z^2)^{\chi(X)} \det(I - zM + z^2(D - I))^{-1}, \tag{2.1}$$

where $\chi(X)$ is the Euler characteristic of X, M is the adjacency matrix (i.e. for vertices u, v, $M(u,v)$ is the number of oriented edges from u to v), and D is the diagonal matrix with entries given by the degrees of the vertices.

Another approach to $\zeta_X(z)$ is through subshifts of finite type, as follows. Each edge in $E(X)$ comes with two orientations and we will write $E(X)^o$ for the set of oriented edges. If $e \in E(X)^o$ then we will write $\mathfrak{o}(e)$ and $\mathfrak{t}(e)$, respectively, for the initial and terminal vertices of e, and $\overline{e}$ for the edge with the opposite orientation. Now consider the space

$$\Sigma(X) = \left\{(e_n)_{n=0}^\infty \in (E(X)^o)^{\mathbb{Z}^+} : \mathfrak{t}(e_n) = \mathfrak{o}(e_{n+1}),\ e_{n+1} \neq \overline{e}_n\ \forall n \ge 0\right\}$$

together with the shift map $\sigma : \Sigma(X) \to \Sigma(X)$ defined by $\sigma((e_n)_{n=0}^\infty) = (e_{n+1})_{n=0}^\infty$. Clearly, $\sigma : \Sigma(X) \to \Sigma(X)$ is a subshift of finite type with state set $\mathcal{S} = E(X)^o$.

The map $\kappa e = \overline{e}$ is a fixed point free involution on $\mathcal{S}$. Furthermore, if A denotes the transition matrix for this subshift then $A(i,j) = 1$ if and only if $A(\kappa j, \kappa i) = 1$. Using this symmetry, it is easy to see that $d_r = d_c$. Moreover, we have

$$d_r = \max_{v\in V(X)} \deg(v) - 1.$$

It is clear that closed geodesics of length n in X correspond exactly to periodic orbits of period n for σ and, hence, $\zeta_X(z) = \zeta_\sigma(z)$. This gives an alternative expression for $\zeta_X(z)$ as a determinant, given by (1.3) above.

Now suppose that we have an infinite graph Y, with vertices $V(Y)$ and edges $E(Y)$, which is a regular cover of X with covering group G. (Since G is a quotient of the fundamental group of X, which is a finitely generated free group, G is automatically finitely generated.) As a natural extension of the finite case, one can define a zeta function

$$\zeta_Y(z) = \prod_{[\gamma]} (1 - z^{|\gamma|})^{-1},$$

where now $[\gamma]$ runs over equivalence classes prime closed geodesics of Y modulo the G-action and $|\gamma|$ denotes the length of any geodesic in $[\gamma]$. (This is the function defined in Definition 2.1 of [**11**]. The situation there is actually a little more general, as they allow covers which are not regular but require that vertex stabilisers are finite.) We claim that this zeta function is equal to $\zeta_{T_\psi}(z)$, for some skew product $T_\psi : \Sigma(X) \times G \to \Sigma(X) \times G$.

The skewing function ψ is defined in the following way. For each vertex $v \in V(X)$, choose a fixed lift $\tilde{v} \in V(Y)$. We then have the following lemma.

LEMMA 2.1. *For each $e \in E(X)^o$, there is a unique $g = g(e) \in G$ such that, if $\tilde{e}$ is a lift of e with $\mathfrak{o}(\tilde{e}) = \widetilde{\mathfrak{o}(e)} \cdot h$, then $\mathfrak{t}(\tilde{e}) = \widetilde{\mathfrak{t}(e)} \cdot gh$.*

PROOF. First, let $\tilde{e}$ be the lift of e with $\mathfrak{o}(\tilde{e}) = \widetilde{\mathfrak{o}(e)}$ and define $g \in G$ by $\mathfrak{t}(\tilde{e}) = \widetilde{\mathfrak{t}(e)} \cdot g$. If $\tilde{e}'$ is another lift of e then it is the translate of $\tilde{e}$ by some $h \in G$ and we have $\mathfrak{o}(\tilde{e}') = \widetilde{\mathfrak{o}(e)} \cdot h$ and $\mathfrak{t}(\tilde{e}') = \widetilde{\mathfrak{t}(e)} \cdot gh$. □

We now define $\psi : \Sigma(X) \to G$ by $\psi((e_n)_{n=0}^\infty) = g(e_0)$.

LEMMA 2.2. *A closed geodesic γ in X, corresponding to a periodic σ-orbit $\tau = \{x, \sigma x, \ldots, \sigma^{n-1}x\}$, lifts to a closed geodesic in Y if and only if $\psi^n(x) = e$.*

PROOF. Suppose $x = (e_n)_{n=0}^\infty$. We will treat the vertex $v := \mathfrak{o}(e_0)$ as the initial point of γ. Let $\tilde{\gamma}$ be the lift of γ that starts at $\tilde{v}$. By Lemma 2.1, $\tilde{\gamma}$ ends at $\tilde{v} \cdot \psi^n(x)$ and is thus closed if and only if $\psi^n(x) = e$. □

We conclude from this that $\zeta_Y(z) = \zeta_{T_\psi}(s)$ and hence the conclusion of Theorem 1.2 holds for $\zeta_Y(z)$, provided G is amenable. More precisely, we have the following.

THEOREM 2.3. *Let X be a finite connected graph and let Y be an infinite regular cover of X with amenable covering group G. Suppose that K_n be a Følner exhaustion of G and let $F \subset Y$ be a fundamental domain for the G-action. Let X_n denote the finite graph*

$$X_n = \bigcup_{g \in K_n} F \cdot g.$$

Then

$$\zeta_Y(z) = \lim_{n \to \infty} \zeta_{X_n}(z)^{1/(\#K_n)},$$

uniformly on compact subsets of $\{z \in \mathbb{C} : |z| < (2\|\mathbb{A}\|)^{-1}\}$.

This result already appeared as Theorem 6.6 of [**11**] except that there the convergence takes place for

$$|z| < \frac{1}{\mathfrak{d} + \sqrt{\mathfrak{d}^2 + 2(\mathfrak{d} - 1)}},$$

where $\mathfrak{d} := \max_{v \in V(X)} \deg(v)$. Since

$$\frac{1}{\mathfrak{d} + \sqrt{\mathfrak{d}^2 + 2(\mathfrak{d} - 1)}} \leq \frac{1}{2\mathfrak{d}} = \frac{1}{2(d_r + 1)} < \frac{1}{2d_r} \leq \frac{1}{2\|\mathbb{A}\|},$$

where we have used $\|\mathbb{A}\| \leq \sqrt{d_r d_c} = d_r$, we see that our approach gives a slightly larger disk of convergence.

Remark 2.4. We end the section by noting two other different approaches to zeta function for infinite graphs. In [**6**], Deitmar considers the following zeta function. Let Y be an infinite graph with edge set $E(Y)$. Let $w : E(Y) \to \mathbb{R}^+$ satisfy $w \in \ell^1(E(Y))$. For a closed geodesic $\gamma = e_0 e_1 \cdots e_{n-1}$, set $w(\gamma) = w(e_0)w(e_1) \cdots w(e_{n-1})$ Then one can define

$$\zeta_{Y,w}(z) = \prod_{\gamma} \left(1 - w(\gamma) z^{|\gamma|}\right)^{-1},$$

where the product is taken over prime closed geodesics. This weighting gives rise to a trace class operator on $\ell^2(E^o)$ and hence to an expression for $\zeta_{Y,w}(z)$ as a determinant (Theorem 1.6 of [**6**]). The special case where w is the indicator function of a finite set $K \subset G$ gives ζ_{σ_K}. It is interesting to ask whether one can obtain convergence results analogous to Theorem 1.2 for as sequence $w_n \in \ell^1(E(Y))$ increasing pointwise to 1 (i.e. for each $e \in E(Y)$, $w_n(e)$ is a sequence increasing to 1).

The recent paper of Lenz, Pogorzelski and Schmidt [**15**] gives a very general approach based on noncommutative geometry.

3. Weighted Zeta Functions and Metric Graphs

In this section, we will introduce generalisations of the zeta functions $\zeta_\sigma(z)$ and $\zeta_{T_\psi}(z)$ which involve a weighting. This will give functions of two variables and we shall show that convergence results similar to Theorem 1.2 hold for these more general objects. As an application, this will give a convergence result for *metric* graphs, analogous to Theorem 2.3.

We start with generalising the zeta function $\zeta_\sigma(z)$ by introducing a weighting. Let $f : \Sigma_A \to \mathbb{R}$ be a strictly positive function satisfying $f(x) = f(x_0, x_1)$. We can then consider a zeta function depending on two complex variables:

$$\zeta_{\sigma,f}(z,s) = \exp \sum_{n=1}^{\infty} \frac{z^n}{n} \sum_{\sigma^n x = x} e^{-s f^n(x)}.$$

(This is just a special case of the generalised zeta functions introduced by Ruelle and much studied in dynamics; see, for example, [**16, 18, 19**].) The series converges to give an analytic function for

$$|z| < \exp(-P(-\mathrm{Re}(s) f, \sigma)),$$

where $P(\cdot, \sigma)$ is the standard pressure function [**16, 19**]. Furthermore, $\zeta_{\sigma,f}(z,s)$ extends to $\mathbb{C}^2$ by the formula

$$\zeta_{\sigma,f}(z,s) = \frac{1}{\det(I - zA_s)},$$

where A_s is the matrix

$$A_s(i,j) = A(i,j) e^{-s f(i,j)}.$$

We may generalise $\zeta_{T_\psi}(z)$ in the same way and define

$$\zeta_{T_\psi,f}(z,s) = \exp \sum_{n=1}^{\infty} \frac{z^n}{n} \sum_{\substack{\sigma^n x = x \\ \psi^n(x)=e}} e^{-sf^n(x)}. \tag{3.1}$$

The domain of convergence of this function may be given in terms of the Gurevič pressure introduced by Sarig [**20**]. If we induce a function $\tilde{f} : \Sigma \times G \to \mathbb{R}$ by $\tilde{f}(x,g) = f(x)$ then its Gurevič pressure $P_{\mathrm{Gur}}(\tilde{f}, T_\psi)$ is defined by

$$P_{\mathrm{Gur}}(\tilde{f}, T_\psi) = \limsup_{n\to\infty} \frac{1}{n} \log \sum_{\substack{\sigma^n x = x \\ \psi^n(x)=e}} e^{f^n(x)}.$$

(This definition does not require f to be positive.) Then $\zeta_{T_\psi,f}(z,s)$ converges to an analytic function in (z,s) for

$$|z| < \exp(-P_{\mathrm{Gur}}(-\mathrm{Re}(s)\tilde{f}, T_\psi)).$$

Let $\mathbb{A}$ be the infinite matrix

$$\mathbb{A}_s((i,g),(j,h)) = \mathbb{A}((i,g),(j,h))e^{-sf(i,j)},$$

which also acts as a bounded operator on $\ell^2(\mathcal{S} \times G)$. (The proof that $\mathbb{A}_s$ is bounded is given below as part of the proof of Lemma 4.1.) If K is a finite subset of G then we can define a finite matrix $A_{K,s}$ and a zeta function $\zeta_{\sigma_{K_n},f}(z,s)$ in the obvious way.

We have the following convergence result, which includes Theorem 1.2 as a special case.

THEOREM 3.1. *Let $\sigma : \Sigma \to \Sigma$ be a subshift of finite type and let $T_\psi : \Sigma \to G \to \Sigma \times G$ be a topologically transitive skew product extension, where G is a finitely generated amenable group. Suppose that K_n, $n \geq 1$ is a Følner exhaustion of G. Then we have*

$$\zeta_{T_\psi,f}(z,s) = \lim_{n\to\infty} \zeta_{\sigma_{K_n},f}(z,s)^{1/(\#K_n)},$$

uniformly on compact subsets of $\{(z,s) \in \mathbb{C}^2 : |z| < (2\|\mathbb{A}_s\|)^{-1}\}$.

As a corollary of this theorem, we have a similar convergence result for the zeta function of a metric graph. Let X be a finite graph, as the previous section, but we now suppose that each edge $e \in E(X)$ is given a positive real length $l(e)$. Thinking of the lengths as a function $l : E(X) \to \mathbb{R}^+$, we call the resulting object a metric graph (X,l). If $\gamma = e_0, \ldots, e_{n-1}$ is a closed geodesic then its length is

$$l(\gamma) = l(e_0) + \cdots + l(e_{n-1})$$

(where the length of an oriented geodesic is the length of the corresponding unoriented edge). We can then define the zeta function

$$\zeta_{(X,l)}(s) = \prod_{\gamma} \left(1 - e^{-sl(\gamma)}\right)^{-1},$$

where again the product is taken over prime closed geodesics. Now let Y be a regular G cover of X, where G is an (infinite) amenable group. The lengths on X lift to Y to give a metric graph (Y,l) and we define the zeta function $\zeta_{(Y,l)}(s)$ by using the prime closed geodesics on Y modulo the G-action. We can write these zeta functions in terms of zeta functions for the shift and the skew product ψ as in the previous section, with the lengths corresponding to a weight f defined by $f(e_0, e_1) = l(e_0)$. We then have the following result.

THEOREM 3.2. *Let (X, l) be a finite connected metric graph and let (Y, l) be an infinite regular cover of (X, l) with amenable covering group G. Suppose that K_n is a Følner exhaustion of G and let $F \subset Y$ be a fundamental domain for the G-action. Let X_n denote the finite graph*

$$X_n = \bigcup_{g \in K_n} F \cdot g.$$

Then

$$\zeta_{(Y,l)}(s) = \lim_{n \to \infty} \zeta_{(X_n,i)}(s)^{1/(\#K_n)},$$

uniformly on compact subsets of $\{s \in \mathbb{C} : \|\mathbb{A}_s\| < 1/2\}$.

4. Traces

Let Tr and det denote the usual trace and determinant of a finite matrix. It is easy to see that

$$\#\{x \in \Sigma : \sigma^n x = x\} = \mathrm{Tr}(A^n)$$

and hence we have the standard results that

$$\begin{aligned}\log \zeta_\sigma(z) &= \sum_{n=1}^{\infty} \frac{z^n}{n} \mathrm{Tr}(A^n) = \sum_{n=1}^{\infty} \frac{1}{n} \sum_{\lambda \in \mathrm{spec}(A)} (\lambda z)^n \\ &= - \sum_{\lambda \in \mathrm{spec}(A)} \log(1 - z\lambda)\end{aligned}$$

and

$$\zeta_\sigma(z) = \frac{1}{\det(I - zA)}.$$

Similar formulae hold for $\zeta_{\sigma_K}(z)$ for any finite subset $K \subset G$.

Consider the Hilbert space

$$H = \ell^2(\mathcal{S} \times G) = \left\{ u : \mathcal{S} \times G \to \mathbb{C} : \sum_{(i,g) \in \mathcal{S} \times G} |u(i,g)|^2 < \infty \right\},$$

with the inner product

$$\langle u, v \rangle = \sum_{(i,g) \in \mathcal{S} \times G} u(i,g)\overline{v(i,g)},$$

and its space of bounded linear operators $B(H)$. We will write $\delta_{(i,g)}$ for the element of H which is equal to one at (i, g) and zero elsewhere.

We need to extend the notion of trace to this setting. Let $T \in B(H)$ be compact. As usual, $|T| = \sqrt{T^*T}$ and T is said to be trace class if

$$\mathrm{Tr}(|T|) := \sum_{(i,g) \in \mathcal{S} \times G} \langle |T| \delta_{(i,g)}, \delta_{(i,g)} \rangle$$

is finite. In this case, we can define the trace of T to be

$$\mathrm{Tr}(T) = \sum_{(i,g) \in \mathcal{S} \times G} \langle T \delta_{(i,g)}, \delta_{(i,g)} \rangle$$

and we have

$$|\mathrm{Tr}(T)| \leq \mathrm{Tr}(|T|). \tag{4.1}$$

Furthermore, we have that the trace class operators form a two-sided ideal in $B(H)$ and that, for $S_1, S_2, T \in B(H)$ with T trace class, we have

$$|\mathrm{Tr}(S_1 T S_2)| \leq \|S_1\| \|S_2\| \mathrm{Tr}(|T|). \tag{4.2}$$

There is a natural representation of $\lambda : G \to B(H)$ given by $\lambda(g)u(i,h) = u(i, g^{-1}h)$. Clearly,

$$\langle \lambda(g)u, \lambda(g)v \rangle = \langle u, v \rangle,$$

for all $g \in G$ and $u, v \in \ell^2(\mathcal{S} \times G)$. Let $\mathcal{A} = \{\lambda(g) : g \in G\}'$, i.e. the elements of $B(H)$ which commute with the representation.

LEMMA 4.1. *For each $s \in \mathbb{C}$, $\mathbb{A}_s \in \mathcal{A}$ and $\|\mathbb{A}_s\| \leq \sqrt{d_r d_c} e^{-\mathrm{Re}(s)\eta(s)}$, where*

$$\eta(s) = \begin{cases} \min\{f(i,j) : A(i,j) = 1\} & \text{if } \mathrm{Re}(s) \geq 0 \\ \max\{f(i,j) : A(i,j) = 1\} & \text{if } \mathrm{Re}(s) < 0. \end{cases}$$

PROOF. For $v \in \ell^2(\mathcal{S} \times G)$,

$$\begin{aligned} \|\mathbb{A}_s v\|_2^2 &= \sum_{(i,g)\in\mathcal{S}\times G} |\mathbb{A}_s v(i,g)|^2 = \sum_{(i,g)\in\mathcal{S}\times G} \left| \sum_{(j,h)\in\mathcal{S}\times G} e^{-sf(i,j)} \mathbb{A}((i,g),(j,h)) v(j,h) \right|^2 \\ &\leq e^{-2\mathrm{Re}(s)\eta(s)} \sum_{(i,g)\in\mathcal{S}\times G} \left| \sum_{(j,h)\in\mathcal{S}\times G} \mathbb{A}((i,g),(j,h)) v(j,h) \right|^2 \\ &= e^{-2\mathrm{Re}(s)\eta(s)} \sum_{(i,g)\in\mathcal{S}\times G} \left| \sum_{j\in\mathcal{S}} A(i,j) v(j, g\psi(i,j)) \right|^2 \\ &\leq e^{-2\mathrm{Re}(s)\eta(s)} \sum_{(i,g)\in\mathcal{S}\times G} \left(\sum_{j\in\mathcal{S}} A(i,j) \right) \left(\sum_{j\in\mathcal{S}} A(i,j) |v(j, g\psi(i,j))|^2 \right) \\ &\leq d_r e^{-2\mathrm{Re}(s)\eta(s)} \sum_{i\in S} \sum_{j\in S} \sum_{g\in G} A(i,j) |v(j, g\psi(i,j))|^2 \\ &\leq d_r d_c e^{-2\mathrm{Re}(s)\eta(s)} \|v\|_2^2, \end{aligned}$$

where we have used that $A(i,j)^2 = A(i,j)$, which shows that $\mathbb{A}_s$ is bounded and gives the estimate on the norm.

Also

$$\begin{aligned} (\mathbb{A}\lambda(g))v(i,h) &= \sum_{(j,h')\in\mathcal{S}\times G} \mathbb{A}((i,h),(j,h'))(\lambda(g)v)(j,h') \\ &= \sum_{(j,h')\in\mathcal{S}\times G} A((i,h),(j,h')) v(j, g^{-1}h') = (\lambda(g)\mathbb{A})v(i,h), \end{aligned}$$

so $\mathbb{A}_s \in \mathcal{A}$. □

The algebra $\mathcal{A}$ admits a finite trace Tr_G defined by

$$\mathrm{Tr}_G(T) = \sum_{i\in S} \langle T\delta_{i,e}, \delta_{i,e} \rangle.$$

For any set $K \subset G$, write $\Pi(K) \in B(H)$ for the orthogonal projection onto the subspace $\ell^2(\mathcal{S} \times K)$ and $\Pi(K)^\perp = \Pi(G \setminus K)$ for the projection onto the orthogonal complement. We note the following simple lemma.

LEMMA 4.2. *Let K be a finite subset of G. Then, for any $T \in \mathcal{A}$, we have*

$$\mathrm{Tr}(\Pi(K)T\Pi(K)) = \#K\mathrm{Tr}_G(T).$$

PROOF. The result follows from a direct calculation. We have,

$$\begin{aligned}
\mathrm{Tr}(\Pi(K)T\Pi(K)) &= \sum_{(i,g)\in S\times G} \left\langle \Pi(K)T\Pi(K)\delta_{(i,g)}, \delta_{(i,g)} \right\rangle \\
&= \sum_{(i,g)\in S\times K} \left\langle \Pi(K)T\delta_{(i,g)}, \delta_{(i,g)} \right\rangle \\
&= \sum_{(i,g)\in S\times K} \left\langle \Pi(K)\left(\sum_{(i',g')\in S\times G} \left\langle T\delta_{(i',g')}, \delta_{(i,g)} \right\rangle \delta_{(i',g')} \right), \delta_{(i,g)} \right\rangle \\
&= \sum_{(i,g)\in S\times K} \left\langle \left(\sum_{(i',g')\in S\times K} \left\langle T\delta_{(i',g')}, \delta_{(i,g)} \right\rangle \delta_{(i',g')} \right), \delta_{(i,g)} \right\rangle \\
&= \sum_{(i,g)\in S\times K} \left\langle T\delta_{(i,g)}, \delta_{(i,g)} \right\rangle \\
&= \sum_{(i,g)\in S\times K} \left\langle T\lambda(g)\delta_{(i,e)}, \lambda(g)\delta_{(i,e)} \right\rangle \\
&= \sum_{(i,g)\in S\times K} \left\langle \lambda(g)T\delta_{(i,e)}, \lambda(g)\delta_{(i,e)} \right\rangle \\
&= \#K \sum_{i\in S} \left\langle T\delta_{(i,e)}, \delta_{(i,e)} \right\rangle = \#K\mathrm{Tr}_G(T),
\end{aligned}$$

where the penultimate line uses that $T \in \mathcal{A}$. □

5. Proof of Theorem 1.2

In this final section, we complete the proof of Theorem 1.2. Let K_n, $n \geq 1$, be a Følner exhaustion of G. By Lemma 4.2, we have

$$\log \zeta_{T_\psi, f}(z, s) = \sum_{k=1}^{\infty} \frac{z^k}{k} \mathrm{Tr}_G(\mathbb{A}_s^k) = \sum_{k=1}^{\infty} \frac{z^k}{k} \frac{\mathrm{Tr}(\Pi(K_n)\mathbb{A}_s^k\Pi(K_n))}{\#K_n},$$

for any $n \geq 1$. We observe that

$$\mathrm{Tr}(A_{K_n,s}^k) = \mathrm{Tr}(\Pi(K_n)\mathbb{A}_s\Pi(K_n))^k)$$

and so we want to estimate

$$\mathrm{Tr}(\Pi(K_n)\mathbb{A}_s^k\Pi(K_n)) - \mathrm{Tr}((\Pi(K_n)\mathbb{A}_s\Pi(K_n))^k).$$

To simplify formulae, write $\Pi_n = \Pi(K_n)$ and $\Pi_n^\perp = \Pi(G \setminus K_n)$. Then, for $k \geq 2$,

$$\begin{aligned}\mathrm{Tr}(\Pi_n \mathbb{A}_s^k \Pi_n) &= \mathrm{Tr}\left(\Pi_n(\mathbb{A}_s(\Pi_n + \Pi_n^\perp))^k \Pi_n\right) \\ &= \mathrm{Tr}((\Pi_n \mathbb{A}_s \Pi_n)^k) + \sum_{\substack{\sigma \in \{\perp, 1\}^{k-1} \\ \sigma \neq (1,1,\ldots,1)}} \mathrm{Tr}\left(\Pi_n \prod_{j=1}^{k-1} (\mathbb{A}_s \Pi_n^{\sigma_j}) \mathbb{A}_s \Pi_n\right).\end{aligned}$$

Consider the terms in the sum on the right-hand side and noting that, since $\sigma \neq (1, 1, \ldots, 1)$, the product

$$\Pi_n \prod_{j=1}^{k-1} (\mathbb{A}_s \Pi_n^{\sigma_j}) \mathbb{A}_s \Pi_n$$

contains at least one term of the form $\Pi_n \mathbb{A}_s \Pi_n^\perp$. Using (4.2), we have the estimate

$$\left| \mathrm{Tr}\left(\Pi_n \prod_{j=1}^{k-1} (\mathbb{A}_s \Pi_n^{\sigma_j}) \mathbb{A}_s \Pi_n\right) \right| \leq \|\mathbb{A}_s\|^{k-1} \mathrm{Tr}(|\Pi_m \mathbb{A}_s \Pi_n^\perp|).$$

Furthermore, from the definition of $\mathbb{A}_s$,

$$\Pi(K_n)\mathbb{A}_s \Pi(G \setminus K_n) = \Pi(K_n)\mathbb{A}_s \Pi(\Omega_n),$$

where

$$\Omega_n \subset \bigcup_{A(i,j)=1} K_n \triangle K_n \psi(i,j),$$

and so

$$\mathrm{Tr}(|\Pi_n \mathbb{A}_s \Pi_n^\perp|) = \mathrm{Tr}(|\Pi_n \mathbb{A}_s \Pi(\Omega_n)|) \leq \|\mathbb{A}_s\| \mathrm{Tr}(\Pi(\Omega_n)) = \|\mathbb{A}_s\| \# \Omega_n.$$

Combining these estimates, we have the following.

LEMMA 5.1. *For any $k \geq 2$ and $n \geq 1$, we have*

$$|\mathrm{Tr}(\Pi(K_n)\mathbb{A}_s^k \Pi(K_n)) - \mathrm{Tr}((\Pi(K_n)\mathbb{A}_s \Pi(K_n))^k)| \leq (2^{k-1} - 1)\|\mathbb{A}_s\|^k \# \Omega_n.$$

We now complete the proof of Theorem 1.2. By Lemma 5.1, we see that

$$\begin{aligned}&\left| \frac{1}{\# K_n} \log \zeta_{\sigma_{K_n}}(z) - \log \zeta_{T_\psi}(z) \right| \\ &= \left| \sum_{k=1}^\infty \frac{z^k}{k} \frac{\mathrm{Tr}((\Pi(K_n)\mathbb{A}_s \Pi(K_n))^k)}{\# K_n} - \sum_{k=1}^\infty \frac{z^k}{k} \frac{\mathrm{Tr}(\Pi(K_n)\mathbb{A}_s^k \Pi(K_n))}{\# K_n} \right| \\ &\leq \left(\sum_{k=1}^\infty \frac{2^{k-1} z^k \|\mathbb{A}_s\|^k}{k} \right) \frac{\# \Omega_n}{\# K_n}.\end{aligned}$$

The series in k converges for

$$|z| < \frac{1}{2\|\mathbb{A}_s\|}$$

and $\lim_{n \to \infty} \# \Omega_n / \# K_n = 0$ by (FE3). Hence we have $\zeta_{\sigma_{K_n}, f}(z, s)^{1/\# K_n}$ converges to $\zeta_{T_\psi, f}(z, s)$ uniformly on compact subsets of this disk. This completes the proof.

References

[1] Hyman Bass, *The Ihara-Selberg zeta function of a tree lattice*, Internat. J. Math. **3** (1992), no. 6, 717–797, DOI 10.1142/S0129167X92000357. MR1194071

[2] G. Chinta, J. Jorgenson, and A. Karlsson, *Heat kernels on regular graphs and generalized Ihara zeta function formulas*, Monatsh. Math. **178** (2015), no. 2, 171–190, DOI 10.1007/s00605-014-0685-4. MR3394421

[3] Bryan Clair, *Zeta functions of graphs with $\mathbb{Z}$ actions*, J. Combin. Theory Ser. B **99** (2009), no. 1, 48–61, DOI 10.1016/j.jctb.2008.04.002. MR2467818

[4] Bryan Clair and Shahriar Mokhtari-Sharghi, *Zeta functions of discrete groups acting on trees*, J. Algebra **237** (2001), no. 2, 591–620, DOI 10.1006/jabr.2000.8600. MR1816705

[5] Bryan Clair and Shahriar Mokhtari-Sharghi, *Convergence of zeta functions of graphs*, Proc. Amer. Math. Soc. **130** (2002), no. 7, 1881–1886, DOI 10.1090/S0002-9939-02-06532-2. MR1896018

[6] Anton Deitmar, *Ihara zeta functions of infinite weighted graphs*, SIAM J. Discrete Math. **29** (2015), no. 4, 2100–2116, DOI 10.1137/140957925. MR3416854

[7] R. Dougall and R. Sharp, *Anosov flows, growth rates on covers and group extensions of subshifts*, arXiv:1904.01423, 2019.

[8] Erling Følner, *On groups with full Banach mean value*, Math. Scand. **3** (1955), 243–254, DOI 10.7146/math.scand.a-10442. MR0079220

[9] Rostislav I. Grigorchuk and Andrzej Żuk, *The Ihara zeta function of infinite graphs, the KNS spectral measure and integrable maps*, Random walks and geometry, Walter de Gruyter, Berlin, 2004, pp. 141–180. MR2087782

[10] Daniele Guido, Tommaso Isola, and Michel L. Lapidus, *Ihara zeta functions for periodic simple graphs*, C^*-algebras and elliptic theory II, Trends Math., Birkhäuser, Basel, 2008, pp. 103–121, DOI 10.1007/978-3-7643-8604-7_5. MR2408138

[11] Daniele Guido, Tommaso Isola, and Michel L. Lapidus, *Ihara's zeta function for periodic graphs and its approximation in the amenable case*, J. Funct. Anal. **255** (2008), no. 6, 1339–1361, DOI 10.1016/j.jfa.2008.07.011. MR2565711

[12] B. Gurevič. Topological entropy for denumerable Markov chains, *Soviet Math. Dokl.* 10, 911–915, 1969.

[13] Harry Kesten, *Full Banach mean values on countable groups*, Math. Scand. **7** (1959), 146–156, DOI 10.7146/math.scand.a-10568. MR0112053

[14] Motoko Kotani and Toshikazu Sunada, *Zeta functions of finite graphs*, J. Math. Sci. Univ. Tokyo **7** (2000), no. 1, 7–25. MR1749978

[15] Daniel Lenz, Felix Pogorzelski, and Marcel Schmidt, *The Ihara zeta function for infinite graphs*, Trans. Amer. Math. Soc. **371** (2019), no. 8, 5687–5729, DOI 10.1090/tran/7508. MR3937307

[16] William Parry and Mark Pollicott, *Zeta functions and the periodic orbit structure of hyperbolic dynamics* (English, with French summary), Astérisque **187-188** (1990), 268. MR1085356

[17] Mark Pollicott and Richard Sharp, *Rates of recurrence for $\mathbf{Z}^q$ and $\mathbf{R}^q$ extensions of subshifts of finite type*, J. London Math. Soc. (2) **49** (1994), no. 2, 401–416, DOI 10.1112/jlms/49.2.401. MR1260121

[18] D. Ruelle, *Generalized zeta-functions for Axiom A basic sets*, Bull. Amer. Math. Soc. **82** (1976), no. 1, 153–156, DOI 10.1090/S0002-9904-1976-14003-7. MR400311

[19] David Ruelle, *Thermodynamic formalism*, Encyclopedia of Mathematics and its Applications, vol. 5, Addison-Wesley Publishing Co., Reading, Mass., 1978. The mathematical structures of classical equilibrium statistical mechanics; With a foreword by Giovanni Gallavotti and Gian-Carlo Rota. MR511655

[20] Omri M. Sarig, *Thermodynamic formalism for countable Markov shifts*, Ergodic Theory Dynam. Systems **19** (1999), no. 6, 1565–1593, DOI 10.1017/S0143385799146820. MR1738951

[21] Audrey Terras, *Zeta functions of graphs*, Cambridge Studies in Advanced Mathematics, vol. 128, Cambridge University Press, Cambridge, 2011. A stroll through the garden. MR2768284

Mathematics Institute, University of Warwick, Coventry CV4 7AL, U.K.
Email address: `R.J.Sharp@warwick.ac.uk`

Contemporary Mathematics
Volume **744**, 2020
https://doi.org/10.1090/conm/744/14988

Invariant measures for Cantor dynamical systems

Sergey Bezuglyi and Olena Karpel

Dedicated to the memory of our friend and colleague Sergiy Kolyada

Abstract. This paper is a survey devoted to the study of probability and infinite ergodic invariant measures for aperiodic homeomorphisms of a Cantor set. We focus mostly on the cases when a homeomorphism has either a unique ergodic invariant measure or finitely many such measures (finitely ergodic homeomorphisms). Since every Cantor dynamical system (X, T) can be realized as a Vershik map acting on the path space of a Bratteli diagram, we use combinatorial methods developed in symbolic dynamics and Bratteli diagrams during the last decade to study the simplex of invariant measures.

Contents

2010 *Mathematics Subject Classification.* Primary 37A05, 37B05; Secondary 28D05, 28C15.
Key words and phrases. ergodic invariant measure, aperiodic homeomorphism, Cantor dynamical system, Bratteli diagram, subshift, uniquely ergodic homeomorphism.
This work is partially supported by the NAS of Ukraine (project "Qualitative, asymptotic and numerical analysis of various classes of differential equations and dynamical systems, their classification, and practical application", No. 0199U192376).

1. Introduction

In this survey, we focus on a classical problem of ergodic theory: for a given dynamical system (X, φ), determine the set $M(X, \varphi)$ of invariant measures (by which we will always mean Borel positive measures in the finite case and Borel positive σ-finite measures in the infinite case). In this generality, the problem is too complicated. To make it more precise, we will consider only aperiodic Cantor dynamical systems, i.e., aperiodic homeomorphisms φ of a Cantor set X. There are many natural examples of such systems including subshifts in symbolic dynamics. We will discuss the significant progress which was made during the last decade in this direction.

Because the problem of finding invariant measures for transformation arises in various areas of mathematics, we hope that this survey may be interesting not only for experts working in the ergodic theory but also for mathematicians who are interested in applications of these results. We included necessary definitions and formulated the most important facts to make this text as much self-contained as possible. We begin with the necessary background.

Let (X, φ) be a topological dynamical system, i.e., φ is a homeomorphism of a compact metric space X. A Borel positive measure μ on X is called *invariant* if $\mu(\varphi(A)) = \mu(A)$ for any Borel set A. By the Kakutani–Markov theorem, such a measure always exists. The set of all probability invariant measures $M(X, \varphi)$ is a Choquet simplex (see [**Phe01**]). Let $E(X, \varphi)$ denote the subset of extreme points of the simplex $M(X, \varphi)$. It is known that this set is formed by ergodic measures for φ. By definition, a measure μ is called *ergodic* if $\varphi(A) = A$ implies that either $\mu(A) = 0$ or $\mu(X \setminus A) = 0$. The cardinality of the set $E(X, \varphi)$ can be either any positive integer, or $\aleph_0$, or continuum. If $|E(X, \varphi)| = 1$, then φ is called *uniquely ergodic*. If $|E(X, \varphi)| = k$, $k \in \mathbb{N}$, then φ is called *finitely ergodic*. The reader can find more information in numerous books on ergodic theory and topological dynamics, we mention in this connection the books [**CFS82**], [**Pet89**], and [**Wal82**].

The question about a complete (or even partial) description of the simplex $M(X, \varphi)$ of invariant probability measures for (X, φ) is one of the most important in ergodic theory. It has a long history and many remarkable results. The cardinality of the set of ergodic measures is an important invariant of dynamical systems. The study of relations between the properties of the simplex $M(X, \varphi)$ and those of the dynamical system (X, φ) is a hard and intriguing problem. There is an extensive list of references regarding this problem, we mention here only the books [**Phe01**], [**Gla03**] and the papers [**Dow91**], [**Dow06**], [**Dow08**] for further citations. In particular, it is important to know conditions under which a system (X, φ) is uniquely ergodic or has a finite number of ergodic measures.

We recall that the simplex $M(X, \varphi)$ plays an important role in classification problems. In particular, it is a complete invariant for orbit equivalence of minimal homeomorphisms of a Cantor set [**GPS95**].

The problem of finding invariant measures of a dynamical system (X, φ) looks rather vague in general setting. There are very few universal results that can be applied to an arbitrary homeomorphism φ. A very productive idea is to replace (X, φ) by an isomorphic model (X_B, φ_B) for which the computation of invariant measures is more transparent. To study invariant measures for a Cantor system (X, φ), we will work with *Bratteli diagrams*, the object that is widely used for constructions of transformation models in various dynamics, see Section 2 for definitions. It is difficult to overestimate the significance of Bratteli diagrams for the theory of dynamical systems. A class of graduated infinite graphs, later called Bratteli diagrams, was originally introduced by Bratteli [**Bra72**] in his breakthrough article on the classification of approximately finite C^*-algebras.

It turned out that the ideas developed by Vershik in ergodic theory [**Ver81**], [**Ver82**] found their application in Cantor dynamics. It was proved in [**HPS92**], that any minimal homeomorphisms φ of a Cantor set X can be represented as a homeomorphism φ_B (called *Vershik map*) acting on the path space X_B of a Bratteli diagram B. The dynamical systems obtained in this way are called *Bratteli–Vershik dynamical systems.* Later on, this approach was realized for non-minimal Cantor dynamical systems [**Med06**] and Borel automorphisms of a standard Borel space [**BDK06**].

The literature devoted to Cantor dynamical systems is very extensive. We do not plan to discuss many interesting directions such as the classification of homeomorphisms up to orbit equivalence, dimension groups, the interplay of Cantor dynamical systems and C^*-algebras, etc. The reader, who is interested in this subject, can be referred to the recent surveys and books [**Ska00**], [**Put10**], [**Dur10**], [**BK16**], [**Put18**] and the research papers [**GPS95**], [**GPS99**], [**GMPS10**], [**DHS99**] (more references can be found in the cited surveys).

The main reason why Bratteli diagrams are convenient to use for the study of homeomorphisms $\varphi : X \to X$ is the fact that various properties of φ become more transparent when one deals with the corresponding Bratteli–Vershik dynamical systems. This observation is related first of all to φ-invariant measures and their supports, to minimal components of φ, structure of φ-orbit, etc. In particular, the study of an ergodic φ-invariant measure μ is reduced, roughly speaking, to the computation of the values of μ on cylinder subsets in the path space of the corresponding Bratteli diagram. In other words, the *structure* of a Bratteli diagram determines completely the invariant measures. In this case we should speak about the invariance with respect to *the tail equivalence relation* because there are Bratteli diagrams that do not admit Vershik maps. We emphasize the difference between simple and non-simple Bratteli diagrams in this context. For an aperiodic homeomorphism φ, the simplex $M(X, \varphi)$ may contain the so called "regular" infinite measures, i.e., the infinite σ-finite measures that take finite (nonzero) values on some clopen sets.

We give one more important observation about Bratteli diagrams. They can be used to construct homeomorphisms of a Cantor set with *prescribed* properties. For instance, it is easy to build a diagram that has exactly k ergodic invariant measures.

A similar picture occurs in symbolic dynamics. Let (X, S) be a subshift, i.e., X is a shift invariant closed subset of $\mathcal{A}^{\mathbb{Z}}$ (in the product topology) where $\mathcal{A}$ is a finite alphabet. Then the combinatorial structure of sequences from the set X can be used to determine invariant measures. Boshernitzan's results about the number of ergodic measures for a minimal subshift give the bounds in terms of the complexity function (see [**Bos84**]). We see a clear similarity between the application of Bratteli diagrams and complexity functions to estimate the number of ergodic measures. This is a reason why we included the recent results extending Boshernitzan's approach (see [**CK19**], [**DF17**], [**DF19**] and the references therein).

In the current paper, we focus on the following problem: how to determine the number of ergodic measures for a given Cantor dynamical system. We distinguish three classes of dynamical systems: uniquely ergodic, finitely ergodic, and "infinitely ergodic" systems. This problem was considered in symbolic dynamics for minimal subshifts by many authors (see the references in Sections 2, 4, and 5).

The outline of the paper is as follows. In Section 2, we give necessary definitions and facts that are used below in the main text. The key concepts are Bratteli diagrams (ordered, simple, non-simple, stationary, finite rank, etc), subshifts, complexity functions. Section 3 contains a description of the simplex of invariant measures in terms of incidence matrices. We also discuss the problem of measure extension from a subdiagram. In other words, the proved results clarify conditions that would guarantee finiteness of measures invariant with respect to the tail equivalence relation. In Section 4, we collected results about uniquely ergodic dynamical systems. These results are formulated either in terms of complexity functions or in terms of Bratteli diagrams. We understand that the variety of uniquely ergodic transformations is vast, and the included results have been chosen to illustrate the discussed methods. In the next section, we consider the results about dynamical systems that have finitely many ergodic measures. For us, the most important sources of examples are stationary and finite rank Bratteli diagrams. In the last section, Setion 6, we consider a class of Bratteli diagrams that have countably many ergodic invariant measures. In the paper, the reader will find a big number of explicit examples to visualize the principal theorems.

2. Basics on Cantor dynamics and Bratteli diagrams

This section contains the basic definitions and facts about topological (in particular, Cantor) dynamical systems. Most of the definitions can be found in the well known books on topological and symbolic dynamics, we refer to [**LM95**], [**Kit98**], [**Køu03**].

2.1. Cantor dynamical systems. A *Cantor set (space)* X is a zero-dimensional compact metric space without isolated points. The topology on X is generated by a countable family of clopen subsets. All such Cantor sets are homeomorphic.

For a homeomorphism $T : X \to X$, denote $Orb_T(x) := \{T^n(x) \mid n \in \mathbb{Z}\}$; the set $Orb_T(x)$ is called the *orbit* of $x \in X$ under the action of T (or simply T-orbit). We consider here mostly *aperiodic* homeomorphisms T, i.e., for every x the set $Orb_T(x)$ is countably infinite.

A homeomorphism $T : X \to X$ is called *minimal* if for every $x \in X$ the orbit $Orb_T(x)$ is dense. Any (aperiodic) homeomorphism T of a Cantor set has a *minimal component*: this is a T-invariant closed non-empty subset Y of X such that $T|_Y$ is minimal on Y.

There are several natural notions of equivalence for Cantor dynamical systems. We give the definitions of conjugacy and orbit equivalence for single homeomorphisms of Cantor sets.

Definition 2.1. Let (X,T) and (Y,S) be two Cantor systems. Then

(1) (X,T) and (Y,S) are *conjugate* (or *isomorphic*) if there exists a homeomorphism $h\colon X \to Y$ such that $h \circ T = S \circ h$.

(2) (X,T) and (Y,S) are *orbit equivalent* if there exists a homeomorphism $h\colon X \to Y$ such that $h(Orb_T(x)) = Orb_S(h(x))$ for every $x \in X$. In other words, there exist functions $n, m\colon X \to \mathbb{Z}$ such that for all $x \in X$, $h \circ T(x) = S^{n(x)} \circ h(x)$ and $h \circ T^{m(x)} = S \circ h(x)$.

Let $\mathcal{B}$ be the Borel σ-algebra generated by clopen subsets of X. We consider only Borel positive measures on $(X, \mathcal{B})$. A measure μ is called probability (finite) if $\mu(X) = 1$ ($\mu(X) < \infty$). Similarly, μ is infinite if $\mu(X) = \infty$. In the latter, we assume that μ is a σ-finite measure. Note, that the infinite σ-finite measures that appear as ergodic invariant measures for aperiodic homeomorphisms of a Cantor set can take finite values on some clopen sets and infinite values on other clopen sets, and these measures are not outer regular (see [**BKMS10**, **BKMS13**, **Kar12a**, **Kar12b**]). Given a Cantor dynamical system (X,T), a Borel measure μ on X is called T-*invariant* if $\mu(TA) = \mu(A)$ for any $A \in \mathcal{B}$. A measure μ is called *ergodic* with respect to T if, for any T-invariant set A, either $\mu(A) = 0$ or $\mu(X \setminus A) = 0$.

Let $M(X,T)$ be the set of all T-invariant probability measures. It is well known that $M(X,T)$ is a Choquet simplex whose extreme points are exactly T-ergodic measures. Denote by $E(X,T)$ the set of extreme points (ergodic measures) in $M(X,T)$. If $M(X,T) = \{\mu\}$, then T is called *uniquely ergodic*. Clearly, in this case $|E(X,T)| = 1$ where $|\cdot|$ denotes the cardinality of a set. If $|E(X,T)| = k < \infty$, we say that (X,T) is *finitely ergodic*.

If two systems, (X,T) and (Y,S), are orbit equivalent, then the corresponding homeomorphism between X and Y (see Definition 2.1) induces the homeomorphism between the sets $M(X,T)$ and $M(Y,S)$ (see for instance [**GPS95**, **GPS99**]).

2.2. Languages on finite alphabets and complexity. In this subsection, we recall some definitions from symbolic dynamics. This material can be found in many books, see e.g. [**LM95**].

Let $\mathcal{A}$ be a finite *alphabet*, then a *word* $w = a_1 \cdots a_k$ in this alphabet is a concatenation of letters a_i in $\mathcal{A}$. The length $|w|$ is the number of letters in w. Let $\mathcal{A}^n$ denote the set of words over $\mathcal{A}$ of length n. Then, by $\mathcal{A}^* = \bigcup_{n=1}^{\infty} \mathcal{A}^n$, we denote the set of all finite nonempty words. It is said that a word $w = a_1 \cdots a_k$ occurs in a word $u = b_1 \cdots b_s$ if $a_1 = b_m, ..., a_k = b_{m+k-1}$. The word w is called a *subword* (or *factor*) of u.

A language $\mathcal{L}$ can be determined in the abstract setting as follows.

Definition 2.2. A set $\mathcal{L}$ of finite words on an alphabet $\mathcal{A}$ is called a *language* if:

(i) $\mathcal{A} \subset \mathcal{L}$,

(ii) for any word w from $\mathcal{L}$, all subwords w' of w belong to $\mathcal{L}$ (the language is *factorial)*;

(iii) for any word $w \in \mathcal{L}$, there exist letters a and b such that $awb \in \mathcal{L}$ (the language is *extendable)*.

Let $\mathcal{L}_n = \mathcal{A}^n \cap \mathcal{L}$ denote the set words in the language $\mathcal{L}$ of length n.

A language $\mathcal{L}$ is called *recurrent* if for any $u, v \in \mathcal{L}$ there exists a word $w \in \mathcal{L}$ such that $uwv \in \mathcal{L}$, and $\mathcal{L}$ is called *uniformly recurrent* if for every $u \in \mathcal{L}$ there exists $m \in \mathbb{N}$ such that u is a subword in every $w \in \mathcal{L}_m$. We consider *aperiodic* languages only ($\mathcal{L}$ is periodic if for every word $w = a_1 \cdots a_{|w|}$ there exists $p \in \mathbb{N}$ such that $a_i = a_{i+p}$ where $1 \leq |w| - p$.

The notion of a language arises naturally in symbolic dynamical systems. We first note that for every infinite sequence $\omega \in \mathcal{A}^{\mathbb{N}}$ of symbols from $\mathcal{A}$, one can define the language $\mathcal{L}(\omega)$ determined by ω as the family of all finite subwords that occur in ω.

One can also define the *language of a subshift* (X, S) where $S : \mathcal{A}^{\mathbb{Z}} \to \mathcal{A}^{\mathbb{Z}}$ denotes the (left) shift, and X is a closed S-subset of $\mathcal{A}^{\mathbb{Z}}$. For a subshift (X, S), we define the language $\mathcal{L}(X)$ of (X, S) as the set of all finite words that occur in the sequences x from X. Clearly, $\mathcal{L}(X)$ is a factorial and extendable language. Conversely, if a language $\mathcal{L}$ on a finite alphabet $\mathcal{A}$ is defined, then there exists a subshift $(X_{\mathcal{L}}, S)$ whose language coincides with $\mathcal{L}$. Indeed, the set $X_{\mathcal{L}}$ is now determined by those sequences from $\mathcal{A}^{\mathbb{Z}}$ whose finite subwords belong to $\mathcal{L}$. It is obvious that

$$\mathcal{L}(X_{\mathcal{L}}) = \mathcal{L}.$$

In other words, the map $(X, S) \to \mathcal{L}(X)$ is a bijection from the set of subshifts to the set of non-empty factorial and extendable languages.

The dynamical properties of subshifts (X, S) can be represented in terms of the corresponding languages. For example, the dynamical system (X, S) is *minimal* if and only if the language $\mathcal{L}(X)$ is uniformly recurrent. In this case, the language $\mathcal{L}(X)$ coincides with $\mathcal{L}(\omega)$ where ω is an arbitrary point (sequence) from X. If one fixes a sequence ω in $\mathcal{A}^{\mathbb{N}}$, then the language $\mathcal{L}(\omega)$ defines a subshift denoted by (X_ω, S).

For every language $\mathcal{L}$, we define the *symbolic complexity function* $p_{\mathcal{L}}(n) : \mathbb{N} \to \mathbb{N}$ by setting

$$p_{\mathcal{L}}(n) = |\mathcal{L}_n|,$$

where $|\cdot|$ here stands for the cardinality of the set of words. If the language $\mathcal{L}$ is defined by a sequence $u \in \mathcal{A}^{\mathbb{Z}}$, then the corresponding complexity function is denoted by $p_u(n)$. Clearly, the complexity function is non-decreasing. Indeed, since the language $\mathcal{L}$ is extendable and factorial, for every word $w \in \mathcal{L}_n$ there exists letter b such that $wb \in \mathcal{L}$. In terms of the corresponding subshift $X_{\mathcal{L}}$, every word w of length n occurring in $\mathcal{L}$ is a part of an infinite sequence belonging to $X_{\mathcal{L}}$, hence there exists at least one word of length $n+1$ which contains w as a subword.

Let (X, S) be a minimal subshift on a finite alphabet $\mathcal{A}$. Then the complexity function $p_X(n)$ can be defined either as that of the corresponding language $\mathcal{L}(X)$ or that of an infinite sequence $\omega \in X$. In both cases, these functions are the same.

The complexity functions have been studied extensively in many papers devoted to languages and symbolic dynamical systems, see, e.g. the survey [**Fer99**] and the bibliography therein. We mention here several results about the complexity functions of dynamical systems.

FACT 2.3. (i) [**Fer96**] Let (X_u, S) and (X_v, S) be symbolic dynamical systems defined by uniformly recurrent sequences u and v from $\mathcal{A}^{\mathbb{N}}$. If (X_u, S) and (X_v, S) are topologically conjugate, then there exists an integer c such that, for all $n > c$,

$$p_u(n - c) \le p_v(n) \le p_u(n + c).$$

(ii) [**CH73**] Let ω be a sequence in $\mathcal{A}^{\mathbb{N}}$. Then ω is ultimately periodic if and only if there exists $n \ge 1$ such that $p_\omega(n) \le n$ if and only if there exists $n \ge 1$ such that $p_\omega(n) = p_\omega(n+1)$.

(iii) [**Pan84**] Let $\zeta : \mathcal{A} \to \mathcal{A}^*$ be a primitive substitution (see Remark 4.5). Then the complexity function $p_u(n)$ of the sequence $u = \zeta(u)$ is sublinear, i.e., there exists C, a positive constant, such that $p_u(n) \le Cn$, for $n \ge 1$. Moreover, the set of differences $p_u(n+1) - p_u(n)$ is bounded [**Cas96**].

(iv) $p_u(m+n) \le p_u(m)p_u(n)$, and the limit $\lim_n n^{-1}\log p_u(n)$ (which therefore exists by Fekete's lemma) is the topological entropy of a sequence.

(v) There are sequences u such that $p_u(n) = n+1$; they are called Sturmian sequences.

2.3. Ordered Bratteli diagrams and Vershik maps. A *Bratteli diagram* is an infinite graph $B = (V, E)$ such that the vertex set $V = \bigcup_{i\ge 0} V_i$ and the edge set $E = \bigcup_{i\ge 0} E_i$ are partitioned into disjoint subsets V_i and E_i where

(i) $V_0 = \{v_0\}$ is a single point;

(ii) V_i and E_i are finite sets, $\forall i \ge 0$;

(iii) there exist $r : E \to V$ (range map r) and $s : E \to V$ (source map s) such that $r(E_i) = V_{i+1}$, $s(E_i) = V_i$, and $s^{-1}(v) \ne \emptyset$, $r^{-1}(v') \ne \emptyset$ for all $v \in V$ and $v' \in V \setminus V_0$.

The set of vertices V_i is called the i-th level of the diagram B. A finite or infinite sequence of edges $(e_i : e_i \in E_i)$ such that $r(e_i) = s(e_{i+1})$ is called a *finite* or *infinite path*, respectively. For $m < n$, $v \in V_m$ and $w \in V_n$, let $E(v,w)$ denote the set of all paths $\overline{e} = (e_1, \dots, e_p)$ with $s(\overline{e}) = s(e_1) = v$ and $r(\overline{e}) = r(e_p) = w$. For a Bratteli diagram B, let X_B be the set of infinite paths starting at the top vertex v_0. We endow X_B with the topology generated by cylinder sets $[\overline{e}]$ where $\overline{e} = (e_0, ..., e_n)$, $n \in \mathbb{N}$, and $[\overline{e}] := \{x \in X_B : x_i = e_i,\ i = 0, \dots, n\}$. With this topology, X_B is a zero-dimensional compact metric space. Indeed, the ultrametric d generating the topology can be given by the following rule: for two different paths $x = (x_i)_{i=1}^\infty$ and $y = (y_i)_{i=1}^\infty$ in X_B, set

$$d(x,y) = \frac{1}{2^n},$$

where $n = \min\{i \in \mathbb{N} : x_i \ne y_i\}$. By assumption, we will consider only such Bratteli diagrams B for which X_B is a *Cantor set*, that is X_B has no isolated points.

Given a Bratteli diagram B, the n-th *incidence matrix* $F_n = (f^{(n)}_{v,w}),\ n \ge 0$, is a $|V_{n+1}| \times |V_n|$ matrix such that $f^{(n)}_{v,w} = |\{e \in E_n : r(e) = v, s(e) = w\}|$ for $v \in V_{n+1}$ and $w \in V_n$. Every vertex $v \in V$ is connected with v_0 by a finite path, and the set of $E(v_0, v)$ of all such paths is finite. If $h^{(n)}_v = |E(v_0, v)|$, then, for all $n \ge 1$, we have

$$h^{(n+1)}_v = \sum_{w \in V_n} f^{(n)}_{v,w} h^{(n)}_w \quad \text{or} \quad h^{(n+1)} = F_n h^{(n)}, \tag{2.1}$$

where $h^{(n)} = (h^{(n)}_w)_{w\in V_n}$. The numbers $h^{(n)}_w$ are usually called *heights* (the terminology comes from using the Kakutani–Rokhlin partitions to build a Bratteli–Vershik system for a homeomorphism of a Cantor space, see below).

We define the following important classes of Bratteli diagrams:

Definition 2.4. Let B be a Bratteli diagram.

(1) We say that B has *finite rank* if for some k, $|V_n| \le k$ for all $n \ge 1$.
(2) We say that a finite rank diagram B has *rank* d if d is the smallest integer such that $|V_n| = d$ infinitely often.

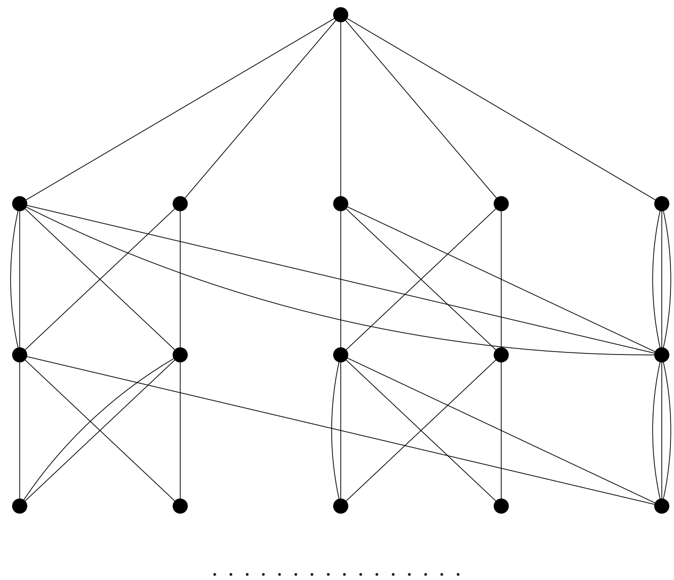


FIGURE 1. Example of a Bratteli diagram

(3) We say that B is *simple* if for any level n there is $m > n$ such that $E(v,w) \neq \emptyset$ for all $v \in V_n$ and $w \in V_m$. Otherwise, B is called *non-simple.*
(4) We say that B is *stationary* if $F_n = F_1$ for all $n \geq 2$.

Let $x = (x_n)$ and $y = (y_n)$ be two paths in X_B. It is said that x and y are *tail equivalent* (in symbols, $(x,y) \in \mathcal{R}$) if there exists some n such that $x_i = y_i$ for all $i \geq n$. Since X_B has no isolated points, the $\mathcal{R}$-orbit of any point $x \in X_B$ is infinitely countable. The diagrams with infinite $\mathcal{R}$-orbits are called *aperiodic.* Note that a Bratteli diagram is simple if the tail equivalence relation $\mathcal{R}$ is minimal.

In order to illustrate the above definitions, we give an example of a nonsimple Bratteli diagram (see Figure 1). This diagram is a non-simple finite rank Bratteli diagram that has exactly two minimal components (they are clearly seen).

We will constantly use the *telescoping* procedure for a Bratteli diagram:

DEFINITION 2.5. Let B be a Bratteli diagram, and $n_0 = 0 < n_1 < n_2 < \ldots$ be a strictly increasing sequence of integers. The *telescoping of B to (n_k)* is the Bratteli diagram B', whose k-level vertex set V'_k is V_{n_k} and whose incidence matrices (F'_k) are defined by

$$F'_k = F_{n_{k+1}-1} \circ \ldots \circ F_{n_k},$$

where (F_n) are the incidence matrices for B.

Roughly speaking, in order to telescope a Bratteli diagram, one takes a subsequence of levels $\{n_k\}$ and considers the set $E(n_k, n_{k+1})$ of all finite paths between the levels $\{n_k\}$ and $\{n_{k+1}\}$ as edges of the new diagram. In particular, a Bratteli diagram B has rank d if and only if there is a telescoping B' of B such that B' has exactly d vertices at each level. When telescoping diagrams, we often do not specify to which levels (n_k) we telescope, because it suffices to know that such a sequence of levels exists.

To avoid consideration of some trivial cases, we will assume that the following *convention* always holds: *our Bratteli diagrams are not unions of two or more disjoint subdiagrams.*

The concept of an ordered Bratteli diagram is crucial for the existence of dynamics on the path space of a Bratteli diagram.

DEFINITION 2.6. A Bratteli diagram $B = (V, E)$ is called *ordered* if a linear order '$>$' is defined on every set $r^{-1}(v)$, $v \in \bigcup_{n\geq 1} V_n$. We denote by ω the corresponding partial order on E and write (B, ω) when we consider B with the ordering ω. Let $\mathcal{O}_B$ denote the set of all orders on B.

Every $\omega \in \mathcal{O}_B$ defines the *lexicographic* order on the set $E(k, l)$ of finite paths between vertices of levels V_k and V_l: $(e_{k+1}, ..., e_l) > (f_{k+1}, ..., f_l)$ if and only if there is i with $k + 1 \leq i \leq l$, such that $e_j = f_j$ for $i < j \leq l$ and $e_i > f_i$. It follows that, given $\omega \in \mathcal{O}_B$, any two paths from $E(v_0, v)$ are comparable with respect to the lexicographic order generated by ω. If two infinite paths are tail equivalent, and agree from the vertex v onwards, then we can compare them by comparing their initial segments in $E(v_0, v)$. Thus, ω defines a partial order on X_B, where two infinite paths are comparable if and only if they are tail equivalent.

DEFINITION 2.7. Let (B, ω) be an ordered Bratteli diagram. We call a finite or infinite path $e = (e_i)$ *maximal (minimal)* if every e_i is maximal (minimal) amongst the edges from the set $r^{-1}(r(e_i))$.

Denote by $X_{\max}(\omega)$ and $X_{\min}(\omega)$ the sets of all maximal and minimal infinite paths in X_B, respectively. It is not hard to see that $X_{\max}(\omega)$ and $X_{\min}(\omega)$ are *non-empty closed subsets* of X_B; in general, $X_{\max}(\omega)$ and $X_{\min}(\omega)$ may have interior points. For a finite rank Bratteli diagram B, the sets $X_{\max}(\omega)$ and $X_{\min}(\omega)$ are always finite for any ω, and if B has rank d, then each of them have at most d elements ([**BKM09**]). For an aperiodic Bratteli diagram B, we see that $X_{\max}(\omega) \cap X_{\min}(\omega) = \emptyset$.

We say that an ordered Bratteli diagram (B, ω) is *properly ordered* if the sets $X_{\max}(\omega)$ and $X_{\min}(\omega)$ are singletons. A Bratteli diagram is called *regular* if the set of maximal paths and set of minimal paths have empty interior.

DEFINITION 2.8. Let (B, ω) be an ordered Bratteli diagram. We say that $\varphi = \varphi_\omega : X_B \to X_B$ is a *(continuous) Vershik map* if it satisfies the following conditions:

(i) φ is a homeomorphism of the Cantor set X_B;

(ii) $\varphi(X_{\max}(\omega)) = X_{\min}(\omega)$;

(iii) if an infinite path $x = (x_0, x_1, \ldots)$ is not in $X_{\max}(\omega)$, then $\varphi(x_0, x_1, \ldots) = (x_0^0, \ldots, x_{k-1}^0, \overline{x_k}, x_{k+1}, x_{k+2}, \ldots)$, where $k = \min\{n \geq 1 : x_n \text{ is not maximal}\}$, $\overline{x_k}$ is the successor of x_k in $r^{-1}(r(x_k))$, and $(x_0^0, \ldots, x_{k-1}^0)$ is the minimal path in $E(v_0, s(\overline{x_k}))$.

If ω is an ordering on B, then one can always define the map φ_0 that maps $X_B \setminus X_{\max}(\omega)$ onto $X_B \setminus X_{\min}(\omega)$ according to (iii) of Definition 2.8. The question about the existence of the Vershik map is equivalent to that of an extension of $\varphi_0 : X_B \setminus X_{\max}(\omega) \to X_B \setminus X_{\min}(\omega)$ to a homeomorphism of the entire set X_B.

If ω is a proper ordering, then φ_ω is a homeomorphism. In particular any simple Bratteli diagram admits a Vershik map.

DEFINITION 2.9. Let B be a Bratteli diagram B. We say that an ordering $\omega \in \mathcal{O}_B$ is *perfect* if ω admits a Vershik map φ_ω on X_B. Denote by $\mathcal{P}_B$ the set of all perfect orderings on B.

We observe that for a regular Bratteli diagram with an order ω, the Vershik map φ_ω, if it exists, is defined in a unique way. Also, a necessary condition for $\omega \in \mathcal{P}_B$ is that $|X_{\max}(\omega)| = |X_{\min}(\omega)|$. Given (B, ω) with $\omega \in \mathcal{P}_B$, the uniquely defined system (X_B, φ_ω) is called a *Bratteli–Vershik* or *adic* system.

We can summarize the above definitions and results in the following statement.

THEOREM 2.10. *Let $B = (V, E, \omega)$ be an ordered Bratteli diagram with a perfect order $\omega \in \mathcal{P}_B$. Then there exists an aperiodic homeomorphism (Vershik map) φ_ω acting on the path space X_B according to Definition* 2.8. *The homeomorphism φ_ω is minimal if and only if B is simple.*

The pair (X_B, φ_ω) is called a *Bratteli–Vershik dynamical system.*

The simplest example of a Bratteli diagram is an *odometer*. Any odometer can be realized as a Bratteli diagram B with $|V_n| = 1$ for all n. Then any order on B is proper and defines the Vershik map.

It is worth noticing that a general Bratteli diagram may have a rather complicated structure. In particular, the tail equivalence relation may have uncountably many minimal components or, in other words, uncountably many simple subdiagrams that do not have connecting edges.

The ideas developed in the papers by Vershik [**Ver81**], [**Ver82**], where sequences of refining measurable partitions of a measure space were used to construct a realization of an ergodic automorphisms of a measure space, turned out to be very fruitful for finding a model of any minimal homeomorphism T of a Cantor set X. In [**HPS92**], Herman, Putnam, and Skau found an explicit construction that allows one to define an ordered simple Bratteli diagram $B = (V, E, \omega)$ such that T is conjugate to the corresponding Vershik map φ_ω. The authors used the existence of the first return time map to any clopen set to build the nested sequence of Kakutani–Rohklin partitions and the corresponding ordered Bratteli diagram. Since this construction is described in many papers (not only in [**HPS92**]), we will not give the details here referring to the original paper and [**Dur10**] for detailed explanation. The case of aperiodic Cantor system is much subtler and was considered in [**BDM05**] and [**Med06**].

Let (X, T) be an aperiodic Cantor system. A closed subset Y of X is called *basic* if (1) $Y \cap T^i Y = \emptyset, i \neq 0$, and (2) every clopen neighborhood A of Y is a complete T-section, i.e., A meets every T-orbit at least once. This means that every point from A is recurrent. It is clear that if T is minimal then every point of X is a basic set. It was proved in [**Med06**] that every aperiodic Cantor system (X, T) has a basic set. This is a crucial step in the proof of the following theorem:

THEOREM 2.11. [**Med06**] *Let (X, T) be a Cantor aperiodic system with a basic set Y. There exists an ordered Bratteli diagram $B = (V, E, \omega)$ such that (X, T) is conjugate to a Bratteli–Vershik dynamical system (X_B, φ_ω). The homeomorphism implementing the conjugacy between T and φ_ω maps the basic set Y onto the set $X_{\min}(\omega)$ of all minimal paths of X_B. The equivalence class of B does not depend on a choice of $\{\xi(n)\}$ with the property $\bigcap_n B(\xi(n)) = Y$.*

Is the converse theorem true? In the case of a *simple Bratteli diagram, the answer is obviously affirmative:* there exists a proper order ω on any simple Bratteli diagram B so that (X_B, φ_ω) is a minimal Cantor system. *For general non-simple Bratteli diagrams the answer is negative.* The first example of a Bratteli diagram that does not admit a Vershik map was found in [**Med06**]. A systematic study of this problem is given in [**BKY14**], [**BY17**], [**JQY17**], see also [**BK16**].

Perfect orderings were also studied in [**DK18, DK19**]. The approach there was a bit different: the starting point was the abstract compact invertible zero-dimensional system (X, T) and the aim was to find an ordered regular Bratteli diagram $B = (V, E, \omega)$ with the perfect ordering ω such that (X_B, φ_ω) is topologically conjugate to (X, T). Regular perfectly ordered Bratteli diagrams are called *decisive*. The following theorem holds.

THEOREM 2.12 ([**DK19**]). *A (compact, invertible) zero-dimensional system (X, T) is topologically conjugate to a decisive Bratteli–Vershik system (X_B, φ_ω) if and only if the set of aperiodic points of (X, T) is dense, or its closure misses one periodic orbit.*

Notice that in the above theorem the space X can have isolated points. The proof uses Krieger's Marker Lemma [**Boy83**] and representation of (X, T) as an array system. Also in [**Shi18**] a non-trivial Bratteli–Vershik model is built for every compact metric zero-dimensional dynamical system.

3. Invariant measures on Bratteli diagrams

Since any aperiodic Cantor dynamical system (X, T) admits a realization as a Bratteli–Vershik dynamical system (see Section 2), the study of T-invariant measures is reduced to the case of measures defined on the path space of a Bratteli diagram. The advantage of this approach is based on the facts that (i) any such a measure is completely determined by its values on cylinder sets of X_B, and (ii) there are simple and explicit formulas for measures of cylinder sets. This method is particularly transparent for stationary and finite rank Bratteli diagrams, simple and non-simple ones [**BKMS10**], [**BKMS13**].

It is worth pointing out that the study of measures on a Bratteli diagram is a more general problem than that in Cantor dynamics. This observation follows from the existence (mentioned above) of Bratteli diagrams that do not support any continuous dynamics on their path spaces which is compatible with the tail equivalence relation. If a Bratteli diagram does not admit a Bratteli–Vershik homeomorphism, then we have to work with the *tail equivalence relation* $\mathcal{R}$ on X_B and study measures invariant with respect to $\mathcal{R}$.

3.1. Simplices, stochastic incidence matrices, examples. In this subsection, we show that the set of all probability invariant measures on a Bratteli diagram corresponds to the inverse limit of a decreasing sequence of convex sets. Let μ be a Borel probability non-atomic $\mathcal{R}$-invariant measure on X_B (for brevity, we will use the term "measure on B" below). We denote the set of all such measures by $\mathcal{M}_1(B)$ and by $\mathcal{E}_1(B)$ the set of all ergodic invariant probability measures. The fact that μ is an $\mathcal{R}$-invariant measure means that $\mu([e]) = \mu([e'])$ for any two cylinder sets $e, e' \in E(v_0, w)$, where $w \in V_n$ is an arbitrary vertex, and $n \geq 1$. Since any measure on X_B is completely determined by its values on clopen (even cylinder) sets, we conclude that in order to define an $\mathcal{R}$-invariant measure μ, one needs to know the sequence of vectors $\overline{p}^{(n)} = (p_w^{(n)} : w \in V_n), n \geq 1$, such that $p_w^{(n)} = \mu([e(v_0, w)])$ where $e(v_0, w)$ is a finite path from $E(v_0, w)$. For $w \in V_n$ it is clear that

$$[e(v_0, w)] = \bigcup_{e(w,v), v \in V_{n+1}} [e(v_0, w), e(w, v)], \tag{3.1}$$

so that $[e(v_0, v)] \subset [e(v_0, w)]$. Relation (3.1) implies that

$$\widetilde{F}_n^T \overline{p}^{(n+1)} = \overline{p}^{(n)}, \ n \geq 1, \tag{3.2}$$

where $\widetilde{F}_n^T$ denotes the transpose of the incidence matrix $\widetilde{F}_n$. The entries of the vectors $\overline{p}^{(n)}$ can be also found by the formula

$$p_w^{(n)} = \frac{\mu(X_w^{(n)})}{h_w^{(n)}},$$

where

$$X_w^{(n)} = \bigcup_{e \in E(v_0, w)} [e], \quad w \in V_n. \tag{3.3}$$

The clopen set $X_w^{(n)}$ is called a *tower*, since it is the tower in the Kakutani–Rokhlin partition that corresponds to the vertex w (see Section 2). The measure of this tower is

$$\mu(X_w^{(n)}) = h_w^{(n)} p_w^{(n)} =: q_w^{(n)}. \tag{3.4}$$

Denote $\overline{q}^{(n)} = (q_w^{(n)} : w \in V_n), n \geq 1$.

Because $\mu(X_B) = 1$, we see that, for any $n > 1$,

$$\sum_{w \in V_n} h_w^{(n)} p_w^{(n)} = \sum_{w \in V_n} q_w^{(n)} = 1.$$

We can obtain a formula similar to (3.2), but for $\overline{q}^{(n)}$ instead of $\overline{p}^{(n)}$ and stochastic incidence matrices F_n instead of usual incidence matrices $\widetilde{F}_n$. The entries of the row stochastic incidence matrix F_n are defined by the formula

$$f_{vw}^{(n)} = \frac{\widetilde{f}_{vw}^{(n)} h_w^{(n)}}{h_v^{(n+1)}}. \tag{3.5}$$

Example 3.1 (Equal row sums (ERS) Bratteli diagrams). In this example, we compute the stochastic incidence matrices for a class of Bratteli diagrams that have the so called *equal row sum (ERS)* property. A Bratteli diagram B has the ERS property if there exists a sequence of natural numbers (r_n) such that the incidence matrices $(\widetilde{F}_n)$ of B satisfy the condition

$$\sum_{w \in V_n} \widetilde{f}_{v,w}^{(n)} = r_n$$

for every $v \in V_{n+1}$. It is known that Bratteli–Vershik systems with the ERS property can serve as models for Toeplitz subshifts (see [**GJ00**]). In particular, we have $\widetilde{F}_0 = \overline{h}^{(1)} = (r_0, \ldots, r_0)^T$. It follows from (2.1) that, for ERS Bratteli diagrams, $h_w^{(n)} = r_0 \cdots r_{n-1}$ for every $w \in V_n$. Hence we have for all $n \geq 1$, $w \in V_n$ and $v \in V_{n+1}$:

$$f_{vw}^{(n)} = \frac{\widetilde{f}_{vw}^{(n)}}{r_n}.$$

In general, it is difficult to compute the elements of the matrix F_n explicitly because the terms $h_w^{(n)}$ used in the formula (3.5) are the entries of the product of matrices. In Section 4, the reader can find Examples 4.14 and 4.16 of Bratteli diagrams, for which stochastic incidence matrices are computed explicitly. Some of the results about the exact number of ergodic invariant measures for a diagram are formulated in terms of $\overline{q}^{(n)}$ and F_n (see Sections 5 and 6). It is easy to prove the following lemma.

LEMMA 3.2. *Let μ be a probability measure on the path space X_B of a Bratteli diagram B. Let (F_n) be a sequence of corresponding stochastic incidence matrices. Then, for every $n \geq 1$, the vector $\overline{q}^{(n)} = (q_v^{(n)} : v \in V_n)$ (see (3.4)) is a probability vector such that*

$$F_n^T \overline{q}^{(n+1)} = \overline{q}^{(n)}, \qquad n \geq 1. \tag{3.6}$$

We see that the formula in (3.6) is a necessary condition for a sequence of vectors $(\overline{q}^{(n)})$ to be defined by an invariant probability measure. It turns out that the converse statement is true, in general. We formulate below Theorem 3.3, where all $\mathcal{R}$-invariant measures are explicitly described.

Using the definition of stochastic incidence matrix F_n (see (3.5)) and Lemma 3.2, we define a decreasing sequence of convex polytopes $\Delta_m^{(n)}$, $n, m \geq 1$, and the limiting convex sets $\Delta_\infty^{(n)}$. They are used to describe the set $\mathcal{M}_1(B)$ of all probability $\mathcal{R}$-invariant measures on B. Namely, denote

$$\Delta^{(n)} := \{(z_w^{(n)})_{w\in V_n}^T : \sum_{w\in V_n} z_w^{(n)} = 1 \text{ and } z_w^{(n)} \geq 0,\ w \in V_n\}.$$

The sets $\Delta^{(n)}$ are standard simplices in the space $\mathbb{R}^{|V_n|}$ with $|V_n|$ vertices $\{\overline{e}^{(n)}(w) : w \in V_n\}$, where $\overline{e}^{(n)}(w) = (0, ..., 0, 1, 0, ...0)^T$ is the standard basis vector, i.e. $e_u^{(n)}(w) = 1$ if and only if $u = w$. Since F_n is a stochastic matrix, we have the obvious property

$$F_n^T(\Delta^{(n+1)}) \subset \Delta^{(n)}, \quad n \in \mathbb{N}.$$

Let μ be a probability $\mathcal{R}$-invariant measure μ on X_B with values $q_w^{(n)}$ on the towers $X_w^{(n)}$. Then $(q_w^{(n)} : w \in V_n)^T$ lies in the simplex $\Delta^{(n)}$. Set

$$\Delta_m^{(n)} = F_n^T \cdot \ldots \cdot F_{n+m-1}^T(\Delta^{(n+m)}) \tag{3.7}$$

for $m = 1, 2, \ldots$ Then we see that

$$\Delta^{(n)} \supset \Delta_1^{(n)} \supset \Delta_2^{(n)} \supset \ldots$$

We write

$$\Delta_\infty^{(n)} = \bigcap_{m=1}^{\infty} \Delta_m^{(n)}. \tag{3.8}$$

It follows from (3.7) and (3.8) that

$$F_n^T(\Delta_\infty^{(n+1)}) = \Delta_\infty^{(n)}, \quad n \geq 1. \tag{3.9}$$

The next theorem describes all $\mathcal{R}$-invariant probability measures on X_B.

THEOREM 3.3 ([**BKMS10**], [**BKK**]). *Let $B = (V, E)$ be a Bratteli diagram with the sequence of stochastic incidence matrices (F_n), and let $\mathcal{M}_1(B)$ denote the set of $\mathcal{R}$-invariant probability measures on the path space X_B.*

(1) If $\mu \in \mathcal{M}_1(B)$, then the probability vector

$$\overline{q}^{(n)} = (\mu(X_w^{(n)}))_{w\in V_n},$$

where $X_w^{(n)}$ is defined in (3.3), satisfies the following conditions for $n \ge 1$:

(i) $\overline{q}^{(n)} \in \Delta_\infty^{(n)}$,

(ii) $F_n^T \overline{q}^{(n+1)} = \overline{q}^{(n)}$.

Conversely, suppose that $\{\overline{q}^{(n)}\}$ is a sequence of non-negative probability vectors such that, for every $\overline{q}^{(n)} = (q_w^{(n)})_{w\in V_n} \in \Delta^{(n)}$ $(n \ge 1)$, the condition (ii) holds. Then the vectors $\overline{q}^{(n)}$ belong to $\Delta_\infty^{(n)}$, $n \in \mathbb{N}$, and there exists a uniquely determined $\mathcal{R}$-invariant probability measure μ such that $\mu(X_w^{(n)}) = q_w^{(n)}$ for $w \in V_n, n \in \mathbb{N}$.

(2) Let Ω be the subset of the infinite product $\prod_{n\ge1} \Delta_\infty^{(n)}$ consisting of sequences $(\overline{q}^{(n)})$ such that $F_n^T \overline{q}^{(n+1)} = \overline{q}^{(n)}$. Then the map

$$\Phi : \mathcal{M}_1(B) \to \Omega$$
$$\mu \mapsto (\overline{q}^{(n)})$$

is an affine isomorphism. Moreover, $\Phi(\mu)$ is an extreme point of Ω if and only if μ is ergodic.

(3) Let B be a Bratteli diagram of rank K. Then the number of ergodic invariant measures on B is bounded above by K and bounded below by the dimension of the finite-dimensional simplex $\Delta_\infty^{(1)}$.

REMARK 3.4. (a) From Theorem 3.3 it follows that the set $\mathcal{M}_1(B)$ can be identified with the inverse limit of the sequence $(F_n^T, \Delta_\infty^{(n)})$. In general, the set $\Delta_\infty^{(n)}$ is a convex subset of the $(|V_n| - 1)$-dimensional simplex $\Delta^{(n)}$. In some cases, which will be considered in Section 5, the set $\Delta_\infty^{(n)}$ is a finite-dimensional simplex itself.

(b) In Theorem 3.3 part (2), the set $\mathcal{M}_1(B)$ can be affinely isomorphic to the set $\Delta_\infty^{(1)}$. For instance, it happens when all stochastic incidence matrices are square non-singular matrices of the same dimension $K \times K$ for some $K \in \mathbb{N}$. This case will be considered in Section 5.

(c) The procedure of telescoping (see Definition 2.5) preserves the set of invariant measures; hence we can apply it when necessary without loss of generality.

In order to find all ergodic invariant measures on a Bratteli diagram, we will study the number of extreme points of $\Delta_\infty^{(n)}$ for every n.

Let

$$G_{(n+m,n)} = F_{n+m} \cdots F_n \tag{3.10}$$

for $m \ge 0$ and $n \ge 1$. Denote the elements of $G_{(n+m,n)}$ by $(g_{uw}^{(n+m,n)})$, where $u \in V_{n+m+1}$ and $w \in V_n$. The sets $\Delta_m^{(n)}, m \ge 0$, defined in (3.7), form a decreasing sequence of convex polytopes in $\Delta^{(n)}$. The vertices of $\Delta_m^{(n)}$ are some (or all) vectors from the set $\{\overline{g}^{(n+m,n)}(v) : v \in V_{n+m+1}\}$, where we denote

$$\overline{g}^{(n+m,n)}(v) = (g_w^{(n+m,n)}(v))_{w\in V_n} = G^T_{(m+n,n)}(\overline{e}^{(n+m+1)}(v)).$$

Obviously, we have the relation

$$\overline{g}^{(n+m,n)}(v) = \sum_{w\in V_n} g_{vw}^{(n+m,n)}\overline{e}^{(n)}(w). \tag{3.11}$$

Let $\{\overline{y}^{(n,m)}(v)\}$ be the set of all vertices of $\Delta_m^{(n)}$. Then $\overline{y}^{(n,m)}(v) = \overline{g}^{(n+m,n)}(v)$ for v belonging to some subset of V_{n+m+1}.

We observe the following fact. Every vector $\overline{q}^{(n)}$ from the set $\Delta_\infty^{(n)}$ can be written in the standard basis as

$$\overline{q}^{(n)} = \sum_{w\in V_n} q_w^{(n)}\overline{e}^{(n)}(w).$$

It turns out that the numbers $q_v^{(n+m+1)}$, $v \in V_{n+m+1}$, are the coefficients in the convex decomposition of $\overline{q}^{(n)}$ with respect to vectors $\overline{g}^{(n+m,n)}(v)$.

Proposition 3.5 ([**BKK**]). *Let $\mu \in \mathcal{M}_1(B)$, and $q_w^{(n)} = \mu(X_w^{(n)})$ ($w \in V_n$) for all $n \in \mathbb{N}$. Then*

$$\overline{q}^{(n)} = \sum_{v\in V_{n+m+1}} q_v^{(n+m+1)}\overline{g}^{(n+m,n)}(v). \tag{3.12}$$

In particular,

$$\overline{q}^{(1)} = \sum_{v\in V_{m+1}} q_v^{(m+1)}\overline{g}^{(m+1,1)}(v).$$

Remark 3.6. For every $n \geq 1$, define

$$\Delta_\infty^{(n),\varepsilon} := \bigcup_{\overline{q}\in\Delta_\infty^{(n)}} B(\overline{q}, \varepsilon),$$

where $B(\overline{q}, \varepsilon)$ is the ball of radius $\varepsilon > 0$ centered at $\overline{q} \in \mathbb{R}^{|V_n|}$. Here the metric is defined by the Euclidean norm $||\cdot||$ on $\mathbb{R}^{|V_n|}$. Fix any natural numbers n and m. Let $\Delta_m^{(n)}$ be defined as above. It can be proved straightforwardly that if $\overline{q}^{(n,m)} \in \Delta_m^{(n)}$ for infinitely many m and $\overline{q}^{(n,m)} \to \overline{q}^{(n)}$ as $m \to \infty$, then $\overline{q}^{(n)} \in \Delta_\infty^{(n)}$. Moreover, for every $\varepsilon > 0$ there exists $m_0 = m_0(n, \varepsilon)$ such that $\Delta_m^{(n)} \subset \Delta_\infty^{(n),\varepsilon}$ for all $m \geq m_0$.

The next statement shows that vertices of the limiting convex set $\Delta_\infty^{(n)}$ can be obtained as limits of sequences of vertices of convex polytopes.

Lemma 3.7 ([**BKK**]). *Fix $n \in \mathbb{N}$ and $\varepsilon > 0$. Let $\Delta_\infty^{(n)}$ and $\Delta_{n+m+1}^{(n)}$, be defined as above for any $m \in \mathbb{N}$. Then, for every vertex $\overline{y} \in \Delta_\infty^{(n)}$ there exists $m_0 = m_0(n, \varepsilon)$ such that, for all $m \geq m_0$, one can find a vertex $\overline{y}^{(n,m)}(v) \in \Delta_{n+m+1}^{(n)}$ satisfying the property*

$$||\overline{y} - \overline{y}^{(n,m)}(v)|| < \varepsilon.$$

3.2. Subdiagrams and measure extension (finite and infinite measures). In this subsection, we study supports of ergodic invariant measures on arbitrary Bratteli diagrams in terms of subdiagrams. By a Bratteli subdiagram, we mean a Bratteli diagram $\overline{B}$ that can be obtained from B by removing some vertices and edges from each level of B. Then $X_{\overline{B}} \subset X_B$. We will consider two extreme cases of Bratteli subdiagrams: vertex subdiagram (when we fix a subset of vertices at each level and take all existing edges between them) and edge subdiagram (some edges are removed from the initial Bratteli diagram but the set of vertices is preserved). It is clear that an arbitrary subdiagram can be obtained as a combination of these cases.

Take a subdiagram $\overline{B}$ and consider the set $X_{\overline{B}}$ of all infinite paths whose edges belong to $\overline{B}$. As a rule, objects related to a subdiagram $\overline{B}$ are denoted by barred symbols. Let $\widehat{X}_{\overline{B}} := \mathcal{R}(X_{\overline{B}})$ be the subset of paths in X_B that are tail equivalent to paths from $X_{\overline{B}}$. Let $\overline{\mu}$ be a probability measure on $X_{\overline{B}}$ invariant with respect to the tail equivalence relation defined on $\overline{B}$. Then $\overline{\mu}$ can be canonically extended to the measure $\widehat{\overline{\mu}}$ on the space $\widehat{X}_{\overline{B}}$ by invariance with respect to $\mathcal{R}$ [**BKMS13**, **ABKK17**]. If we want to extend $\widehat{\overline{\mu}}$ to the whole space X_B, we set $\widehat{\overline{\mu}}(X_B \setminus \widehat{X}_{\overline{B}}) = 0$.

This subsection is devoted to answering the following questions:

(A) Given a subdiagram $\overline{B}$ of B and an ergodic measure μ on X_B, under what conditions on $\overline{B}$ does the subset $X_{\overline{B}}$ have positive measure μ in X_B?

(B) Let ν be a measure supported by the path space $X_{\overline{B}}$ of a subdiagram $\overline{B} \subset B$. Then ν is extended to the subset $\mathcal{R}(X_{\overline{B}})$ by invariance with respect to the tail equivalence relation $\mathcal{R}$. Under what conditions is $\nu(\mathcal{R}(X_{\overline{B}}))$ finite (or infinite)?

In this subsection, we keep the following notation: $\overline{X}_v^{(n)}$ stands for the tower in a subdiagram $\overline{B}$ that is determined by a vertex v of $\overline{B}$. Thus, we consider the paths in $\overline{X}_v^{(n)}$ that contain edges from $\overline{B}$ only. Let $\overline{h}_v^{(n)}$ be the height of the tower $\overline{X}_v^{(n)}$. The following theorem gives criteria for finiteness of the measure extension.

Theorem 3.8 ([**BKK15**]). *Let B be a Bratteli diagram with the sequence of incidence matrices $\{\widetilde{F}_n\}_{n=0}^\infty$ and corresponding stochastic matrices $\{F_n\}_{n=0}^\infty$. Let $\overline{B}$ be a vertex subdiagram of B defined by the sequence of subsets $\{W_n\}_{n=0}^\infty$, $W_n \subset V_n$. Suppose that $\overline{\mu}$ is a probability $\overline{\mathcal{R}}$-invariant measure on $X_{\overline{B}}$. Then the following properties are equivalent:*

$$
\begin{aligned}
\widehat{\overline{\mu}}(\widehat{X}_{\overline{B}}) < \infty \quad &\Longleftrightarrow \quad \sum_{n=1}^{\infty} \sum_{v \in W_{n+1}} \sum_{w \notin W_n} \widetilde{f}_{v,w}^{(n)} h_w^{(n)} \overline{p}_v^{(n+1)} < \infty \\
&\Longleftrightarrow \quad \sum_{n=1}^{\infty} \sum_{v \in W_{n+1}} \widehat{\overline{\mu}}(X_v^{(n+1)}) \sum_{w \notin W_n} f_{v,w}^{(n)} < \infty \\
&\Longleftrightarrow \quad \sum_{i=1}^{\infty} \left(\sum_{w \in W_{i+1}} h_w^{(i+1)} \overline{p}_w^{(i+1)} - \sum_{w \in W_i} h_w^{(i)} \overline{p}_w^{(i)} \right) < \infty.
\end{aligned}
$$

The analogue of Theorem 3.8 can be proved also for edge subdiagrams (see [**ABKK17**]). The following proposition gives a sufficient condition of the finiteness of the measure extension (more necessary and sufficient conditions can be found in [**BKK15**, **ABKK17**]).

Proposition 3.9 ([**BKK15**]). *Let B be a Bratteli diagram with the sequence of stochastic incidence matrices $\{F_n\}_{n=0}^{\infty}$, and let $\overline{B}$ be its subdiagram defined by a sequence of vertices W_n. If*

$$\sum_{n=1}^{\infty} \max_{v \in W_{n+1}} \left(\sum_{w \notin W_n} f_{vw}^{(n)} \right) < \infty,$$

then any tail invariant probability measure $\overline{\mu}$ on $X_{\overline{B}}$ extends to a finite invariant measure $\widehat{\overline{\mu}}$ on $\widehat{X}_{\overline{B}}$.

The following theorem gives a necessary and sufficient condition for a subdiagram $\overline{B}$ of B to have a path space of zero measure in X_B. Though the theorem is formulated for a vertex subdiagram, the statement remains true also for any edge subdiagram $\overline{B}$.

Theorem 3.10 ([**ABKK17**]). *Let B be a simple Bratteli diagram, and let μ be any ergodic probability measure on X_B. Suppose that $\overline{B}$ is a vertex subdiagram of B defined by a sequence (W_n) of subsets of V_n. Then $\mu(X_{\overline{B}}) = 0$ if and only if for all $\varepsilon > 0$ there exists $n = n(\varepsilon)$ such that for all $w \in W_n$ one has*

$$\frac{\overline{h}_w^{(n)}}{h_w^{(n)}} < \varepsilon. \tag{3.13}$$

In fact, Theorem 3.10 states that if a subdiagram $\overline{B}$ satisfies (3.13), then $X_{\overline{B}}$ has measure zero with respect to every ergodic invariant measure, that is the set $X_{\overline{B}}$ is *thin* according to the definition from [**GPS04**]. The following result is a corollary of Theorem 3.10:

Theorem 3.11 ([**ABKK17**]). *Let $\overline{B}$ be a subdiagram of B such that $X_{\overline{B}}$ is a thin subset of X_B. Then for any probability invariant measure $\overline{\mu}$ on $\overline{B}$ we have $\widehat{\overline{\mu}}(\widehat{X}_{\overline{B}}) = \infty$.*

Remark 3.12. There are many papers, where invariant measures for various Bratteli diagrams are studied. For instance, in [**FP08**, **PV10**] the authors consider ergodic invariant probability measures on a Bratteli diagram of a special form, called an Euler graph; the combinatorial properties of the Euler graph are connected to those of Eulerian numbers. The authors of [**FO13**] study spaces of invariant measures for a class of dynamical systems which is called polynomial odometers. These are adic maps on regularly structured Bratteli diagrams and include the Pascal and Stirling adic maps as examples. Petersen [**Pet12**] considers ergodic invariant measures on a Bratteli–Vershik dynamical system, which is based on a diagram whose path counts from the root are the Delannoy numbers.

We would like to mention also the interesting paper by Fisher [**Fis09**] where various properties of Bratteli diagrams and measures are discussed.

4. Uniquely ergodic Cantor dynamical systems

In this section, we consider a number of results about uniquely ergodic Cantor dynamical systems. We are not trying to mention all existing classes of uniquely ergodic homeomorphisms. In the case of symbolic systems, we discuss the results related mostly to the complexity function (Subsection 4.1). More general approach using Bratteli diagrams is considered in Subsection 4.2.

4.1. Minimal uniquely ergodic homeomorphisms in symbolic dynamics. In this subsection, we partially use some statements formulated and proved in [**FM10**] and in [**Bos84**], [**Bos85**], and [**Bos92**].

Definition 4.1. Let $\omega = (\omega_i)$ be an infinite sequence in $\mathcal{A}^{\mathbb{N}}$. It is said that the infinite sequence ω has *uniform frequencies* if, for every factor w of ω,

$$\frac{|\omega_k \cdots \omega_{k+n}|_w}{n+1} \to f_w(\omega) \quad (n \to \infty)$$

uniformly in k. (Here $|u|_w$ denotes the number of occurrences of the word w in the word u)

The following result follows immediately from the ergodic theorem.

Fact 4.2 (Folklore). (i) Let (X, S) be a subshift, and let μ be an S-invariant ergodic measure. Then, for μ-a.e. $\omega \in X$ and for any finite word w in $\mathcal{L}(X)$, the frequency $f_w(\omega)$ exists and is equal to $\mu([w])$ where $[w]$ denotes the corresponding cylinder subset of X.

(ii) A subshift (X_ω, S) is uniquely ergodic if, and only if, the sequence ω has uniform frequencies.

As one of our goals is to discuss relations between the complexity functions and the number of ergodic measures, we recall Boshernitzan's results about uniquely ergodic subshifts. In fact, Boshernitzan proved several impressive results on the cardinality of the set of ergodic measures for minimal subshifts which are based on a careful study of the growth of the complexity functions. The case of finite ergodicity is considered below in Section 5.

Theorem 4.3 ([**Bos84**]). *Let $p_X(n)$ denote the complexity function of a minimal subshift (X, S) over a finite alphabet. If either*

$$\limsup_{n\to\infty} \frac{p_X(n)}{n} < 3,$$

or

$$\liminf_{n\to\infty} \frac{p_X(n)}{n} = \alpha < 2,$$

then (X, S) is uniquely ergodic.

Let X be a compact metric space, $\mathcal{B}$ the Borel σ-algebra, and μ a Borel probability measure on $\mathcal{B}$. Suppose $T : X \to X$ is a measurable map preserving the measure μ. A point $x \in X$ is called a *generic point* for the measure μ if for every continuous function $f : X \to \mathbb{R}$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) = \int_X f \, d\mu.$$

The measure μ is *generic* if it has a generic point.

By the pointwise ergodic theorem, if μ is ergodic, then almost every point with respect to μ is generic. However, there are generic measures which are not ergodic. An example of such a measure is given in [**CM15**] where an interval exchange transformation has a generic non-ergodic measure.

Theorem 4.4. [**Gla03**] *Let (X, T) be a minimal system. Then (X, T) is uniquely ergodic if and only if every point x in X is generic for some measure in $M(X, T)$.*

REMARK 4.5. In this remark we point out several classes of uniquely ergodic Cantor dynamical systems.

(1) *Substitution dynamical systems.* Let $\mathcal{A}$ be a finite alphabet, and let $\tau : \mathcal{A} \to \mathcal{A}^*$ be a substitution. The language of the substitution consists of all words which are seen in $\bigcup_{n=0}^{\infty} \tau^n(\mathcal{A})$. The corresponding subshift (X_τ, S) is called a *substitution dynamical system.* The literature on these dynamical systems is very extensive, we refer to [**Que10**], [**Fog02**]. A substitution τ is called *primitive* if for any $a, b \in \mathcal{A}$ there exists $n \in \mathbb{N}$ such that $|\tau^n(b)|_a \geq 1$. The corresponding substitution dynamical system is uniquely ergodic.

In Section 5 below, we consider *aperiodic* (non-minimal) substitution systems. The situation with invariant measures is different. They may have finitely many egodic invariant probability measures and finitely many infinite ergodic invariant measures as well.

(2) *Linearly recurrent dynamical systems.* Let $\mathcal{A}$ be an alphabet, and let ω be a sequence from $\mathcal{A}^{\mathbb{N}}$ with the language $\mathcal{L}(\omega)$. For a word $u \in \mathcal{L}(\omega)$, we call a word w a return word to u in ω if wu belongs to $\mathcal{L}(\omega)$, u is a prefix of wu, and u has exactly two occurrences in wu (we follow [**Dur98**], more general approach to the notion of return words is given in [**Dur10**]). Denote by $\mathcal{R}_{\omega,u}$ the set of return words to u of ω. When ω is a uniformly recurrent sequence from $\mathcal{A}^{\mathbb{N}}$, then the set $\mathcal{R}_{\omega,u}$ is finite for all $u \in \mathcal{L}(\omega)$.

It is said that a sequence ω is *linearly recurrent* (with constant $K \in \mathbb{N}$) if it is uniformly recurrent and if for all $u \in \mathcal{L}(\omega)$ and all $w \in \mathcal{R}_{\omega,u}$ we have $|w| \leq K|u|$. A subshift (X, S) is called *linearly recurrent* (with constant K) if it is minimal and contains a linearly recurrent sequence (with constant K). In fact, for any $x, y \in X$, we have $\mathcal{R}_{x,u} = \mathcal{R}_{y,u}$.

As proved in [**Dur00**] (see also [**Dur10**]), linearly recurrent subshifts are uniquely ergodic. This result can be deduced from [**Bos92**] or proved directly. One more important fact that relates linearly recurrent subshifts and Bratteli diagrams is proved in [**Dur10**]). It states that such subshifts have an expansive Bratteli–Vershik representation whose incidence matrices belong to a finite set.

4.2. Finite rank Bratteli diagrams and general case. In this subsection, we discuss the results on unique ergodicity of Bratteli diagrams. It is worth recalling that these results describe Cantor dynamical systems which are represented by the corresponding Bratteli diagrams. We give a criterion and sufficient conditions for the unique ergodicity of a Bratteli diagram B of arbitrary rank, in other words, we discuss the case when the space $\mathcal{M}_1(B)$ is a singleton.

We first begin with a class of Bratteli diagrams that have an exact finite rank.

DEFINITION 4.6. It is said that a finite rank Bratteli diagram has an *exact finite rank* if there is a finite invariant measure μ and a constant $\delta > 0$ such that after a telescoping $\mu(X_w^{(n)}) \geq \delta$ for all levels n and vertices w.

The following result shows that the Vershik map on the path space of an exact finite rank diagram cannot be strongly mixing independently of the ordering.

THEOREM 4.7 ([**BKMS13**]). *Let $B = (V, E, \omega)$ be an ordered simple Bratteli diagram of exact finite rank and μ is as in Definition* 4.6.
(1) The diagram B is uniquely ergodic and μ is the unique invariant measure.
(2) Let $\varphi_\omega : X_B \to X_B$ be the Vershik map defined by the order ω on B (φ_ω is not

necessarily continuous everywhere). Then the dynamical system $(X_B, \mu, \varphi_\omega)$ *is not strongly mixing with respect to the unique invariant measure* μ.

On the other hand, it is proved in the same paper that for the so-called "left- to-right" ordering, the Vershik map is not strongly mixing on all finite rank diagrams.

REMARK 4.8. Theorem 4.7 can be viewed as an analogue of a result from [**Bos92**]. To be more precise, let (X, S) be a minimal subshift on a finite alphabet, and let μ be a probability S-invariant measure. Set

$$\varepsilon(n) = \min\{\mu([w]) : w \in \mathcal{L}_n(X)\},$$

where $[w]$ is the cylinder subset of X defined by the word w. If

$$\lim_{n\to\infty} n\varepsilon(n) = 0,$$

then the subshift (X, S) is not uniquely ergodic.

In what follows, we focus on the following problem: find conditions on the (stochastic) incidence matrices ensuring that the diagram is uniquely ergodic.

THEOREM 4.9 ([**BKK**]). *A Bratteli diagram* $B = (V, E)$ *is uniquely ergodic if and only if there exists a telescoping* B' *of* B *such that*

$$\lim_{n\to\infty} \max_{v,v' \in V_{n+1}} \left(\sum_{w\in V_n} \left| f_{vw}^{(n)} - f_{v'w}^{(n)} \right| \right) = 0, \tag{4.1}$$

where $f_{vw}^{(n)}$ *are the entries of the stochastic matrix* F_n *defined by the diagram* B'.

The proof is based on the representation of an invariant measure as a point of the inverse limit of the sequence $(F_n^T, \Delta_\infty^{(n)})$ (see Section 3). We use the fact that B is uniquely ergodic if and only if the set $\Delta_\infty^{(n)}$ is a singleton for all $n = 1, 2, \dots$ and that the polytope $\Delta_m^{(n)}$ is the convex hull of the vectors $\{\overline{g}^{(n+m,n)}(v)\}_{v\in V_{n+m+1}}$ for all $m \in \mathbb{N}$.

The following statement is a corollary of Theorem 4.9 and provides a sufficient condition for a Bratteli diagram to be uniquely ergodic. Note that this condition does not require telescoping.

THEOREM 4.10 ([**BKK**]). *Let* B *be a Bratteli diagram of arbitrary rank with stochastic incidence matrices* F_n *and let*

$$m_n = \min_{v\in V_{n+1}, w\in V_n} f_{vw}^{(n)}.$$

If

$$\sum_{n=1}^{\infty} m_n = \infty, \tag{4.2}$$

then B *is uniquely ergodic.*

A number of sufficient conditions for unique ergodicity of a finite rank Bratteli–Vershik system are obtained in [**BKMS13**]. Here we present some of them.

DEFINITION 4.11. (see e.g. [**Har02**]) (i) For two positive vectors $x, y \in \mathbb{R}^d$, the *projective metric* is defined by the formula

$$D(x,y) = \ln \max_{i,j} \frac{x_i y_j}{x_j y_i} = \ln \frac{\max_i \frac{x_i}{y_i}}{\min_j \frac{x_j}{y_j}},$$

where (x_i) and (y_i) are entries of the vectors x and y.

(ii) For a non-negative matrix A, the *Birkhoff contraction coefficient* is

$$\tau(A) = \sup_{x,y>0} \frac{D(Ax, Ay)}{D(x,y)}.$$

THEOREM 4.12 ([**BKMS13**]). *Let B be a simple Bratteli diagram of finite rank with incidence matrices $\{\widetilde{F}_n\}_{n\geq 1}$. Let $\widetilde{A}_n = \widetilde{F}_n^T$. Then the diagram B is uniquely ergodic if and only if*

$$\lim_{n\to\infty} \tau(\widetilde{A}_m \dots \widetilde{A}_n) = 0 \textit{ for every } m.$$

For a positive matrix $A = (a_{i,j})$, let

$$\phi(A) = \min_{i,j,r,s} \frac{a_{i,j} a_{r,s}}{a_{r,j} a_{i,s}}.$$

If A has a zero entry, then, by definition, we set $\phi(A) = 0$. As noticed in [**Har02**],

$$\tau(A) = \frac{1 - \sqrt{\phi(A)}}{1 + \sqrt{\phi(A)}}$$

in case A has a nonzero entry in each row.

The following result gives sufficient conditions for unique ergodicity which can be easily verified when a diagram is given by a sequence of incidence matrices.

PROPOSITION 4.13 ([**BKMS13**]). *Let $\{\widetilde{A}_n\}_{n\geq 1} = \widetilde{F}_n^T$ be primitive incidence matrices of a finite rank diagram B.*

(1) If

$$\sum_{n=1}^{\infty} \sqrt{\phi(A_n)} = \infty,$$

then B admits a unique invariant probability measure.

(2) If

$$\sum_{n=1}^{\infty} \left(\frac{\widetilde{m}_n}{\widetilde{M}_n} \right) = \infty,$$

where $\widetilde{m}_n$ and $\widetilde{M}_n$ are the smallest and the largest entry of $\widetilde{A}_n$ respectively, then B admits a unique invariant probability measure.

(3) Let $||A||_1 := \sum_{i,j} |a_{i,j}|$. If $||\widetilde{F}_n||_1 \leq Cn$ for some $C > 0$ and all sufficiently large n, then the diagram admits a unique invariant probability measure. In particular, this result holds if the diagram has only finitely many different incidence matrices.

4.3. Examples. The following examples illustrate the results of Subsection 4.2. In particular, Examples 4.14 and 4.16 show that telescoping and using the stochastic incidence matrix are crucial for Theorem 4.9.

EXAMPLE 4.14. Let B be a Bratteli diagram with incidence matrices

$$\widetilde{F}_n = \begin{pmatrix} n & 1 \\ 1 & n \end{pmatrix}, \qquad n \in \mathbb{N}.$$

Then the diagram B is uniquely ergodic, see details in [**BKMS13**], [**ABKK17**, Example 3.6], and [**FFT09**].

Notice that B has the ERS property (see Example 3.1). Hence the corresponding stochastic incidence matrices are:

$$F_n = \begin{pmatrix} 1 - \dfrac{1}{n+1} & \dfrac{1}{n+1} \\ \\ \dfrac{1}{n+1} & 1 - \dfrac{1}{n+1} \end{pmatrix}.$$

Obviously, without telescoping, for B the limit in (4.1) equals 2. However, the telescoping procedure reveals that the diagram is in fact uniquely ergodic.

Suppose we have an ERS diagram with 2×2 stochastic incidence matrices

$$F'_n = \begin{pmatrix} a_n & b_n \\ b_n & a_n \end{pmatrix}.$$

As before, let $G_{(n,n+m)} = (g_{vw}^{(n+m,n)})$ be the corresponding product matrix. It can be easily proved by induction that, for arbitrary $n, m \in \mathbb{N}$, the following formula holds:

$$S(n,m) = \sum_{w \in V_n} \left| g_{vw}^{(n+m,n)} - g_{v'w}^{(n+m,n)} \right| = 2 \prod_{i=0}^{m} |(a_{n+i} - b_{n+i})| .$$

In the case of the diagram B, we obtain

$$S(n,m) = 2 \prod_{i=0}^{m} \left(1 - \frac{2}{n+i+1}\right), \quad n, m \in \mathbb{N}.$$

Since the harmonic series $\sum_{n=1}^{\infty} n^{-1}$ diverges, we see that

$$S(n,m) \to 0, \qquad m \to \infty.$$

Choose a decreasing sequence (ε_k) such that $\varepsilon_k \to 0$ as $k \to \infty$. For $n = n_1$ and ε_1, find m_1 such that $S(n_1, m_1) < \varepsilon_1$. Set $n_2 = n_1 + m_1$. For ε_2, find m_2 such that $S(n_2, m_2) < \varepsilon_2$. Set $n_3 = n_2 + m_2$. Continuing in the same manner, we construct a sequence (n_k) such that $S(n_k, n_{k+1} - n_k) < \varepsilon_k$. Telescope the diagram with respect to the levels (n_k). By Theorem 4.9, we conclude that the diagram B is uniquely ergodic. Notice that the diagram B also satisfies the sufficient condition of unique ergodicity (4.2).

EXAMPLE 4.15. Let B be a simple Bratteli diagram with incidence matrices

$$\widetilde{F}_n = \begin{pmatrix} f_1^{(n)} & 1 & \cdots & 1 \\ 1 & f_2^{(n)} & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & f_d^{(n)} \end{pmatrix}.$$

Let $q_n = \max\{\widetilde{f}_i^{(n)}\widetilde{f}_j^{(n)} : i \neq j\}$. By Proposition 4.13, if for $\widetilde{A}_n = \widetilde{F}_n^T$

$$\sum_{n=1}^{\infty} \sqrt{\phi(\widetilde{A}_n)} = \sum_{n=1}^{\infty} \frac{1}{\sqrt{q}_n} = \infty,$$

then there is a unique invariant probability measure on B.

EXAMPLE 4.16. Let B be the stationary non-simple Bratteli diagram defined by the incidence matrices

$$\widetilde{F}_n = \widetilde{F} = \begin{pmatrix} 3 & 0 \\ 1 & 2 \end{pmatrix}$$

for every $n \in \mathbb{N}$. It is well known that B has a unique finite ergodic measure supported by the 3-odometer (see e.g. [**BKMS10**]). We show that B satisfies the condition of unique ergodicity formulated in Theorem 4.9. It is easy to check that the n-th power of $\widetilde{F}$ is

$$\widetilde{F}^n = \begin{pmatrix} 3^n & 0 \\ 3^n - 2^n & 2^n \end{pmatrix}.$$

Hence the entries of the matrix $\widetilde{F}$ do not satisfy (4.1) even after taking products of these matrices (which corresponds to telescoping of B). Notice that B has the ERS property. For any $n \in \mathbb{N}$ and a vertex $w \in V_n$, we have $h_w^{(n)} = 3^n$. Therefore, the corresponding stochastic incidence matrix and its n-th power are

$$F = \begin{pmatrix} 1 & 0 \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}$$

and

$$F^n = \begin{pmatrix} 1 & 0 \\ 1 - \dfrac{2^n}{3^n} & \dfrac{2^n}{3^n} \end{pmatrix}.$$

Hence, we see that B satisfies (4.1) and is uniquely ergodic. Note that B does not satisfy the sufficient condition of unique ergodicity (4.2) and that Theorem 4.12 and Proposition 4.13 are not applicable since B is not simple.

5. Finitely ergodic Cantor dynamical systems

This section is mostly devoted to the study of aperiodic Cantor dynamical systems which can be represented by Bratteli diagrams with uniformly bounded number of vertices on each level. It is an open question which classes of Cantor dynamical systems admit such a representation.

5.1. Finitely ergodic subshifts. As in Section 4, we begin with the case of finitely ergodic (minimal) subshifts. The recent progress made in [**CK19**], [**DF17**], [**DF19**] essentially improved the known results on the bounds of the cardinality of the set $E(X, S)$ of ergodic invariant measures.

In [**Bos84**], the following remarkable results were proved.

THEOREM 5.1 ([**Bos84**]). *Let (X, S) be a minimal subshift on a finite alphabet $\mathcal{A}$.*

(i) If

$$\liminf_{n\to\infty} \frac{p_X(n)}{n} = \alpha,$$

then $|E(X,S)| \le [\alpha]$, *where* $[\alpha]$ *is the integer part of* α.
(ii) If

$$\limsup_{n\to\infty} \frac{p_X(n)}{n} = \alpha$$

and $\alpha \ge 2$, *then* $|E(X,S)| \le [\alpha] - 1$.

The results given in [**FM10**] extend Boshernitzan's bounds to the so called K-deconnectable symbolic systems. We cite only one of these results from that paper here.

THEOREM 5.2 ([**Mon09**]). *Let* $K \ge 3$ *be an integer. A minimal symbolic system* (X,S) *such that*

$$\limsup_{n\to\infty} \frac{p_X(n)}{n} < K$$

admits at most $K-2$ *ergodic invariant measures.*

In [**DF17**], the authors continued this line of study of the set $E(X,S)$ and considered complexity functions with eventually constant growth condition. By definition, this means that the complexity function $p_X(n)$ of a minimal subshift satisfies the condition: for some $K \in \mathbb{N}$ and all $n \ge n_0$

$$p_X(n+1) - p_X(n) = K. \tag{5.1}$$

Equivalently, $p_X(n) = Kn + C$ for all $n \ge n_0$ for a constant $C \in \mathbb{N}_0$.

THEOREM 5.3 ([**DF17**]). *If the complexity function of a minimal subshift* (X,S) *satisfies eventually constant growth condition (*5.1*) with* $K \ge 4$, *then* $|E(X,S)| \le K-2$.

In the very recent paper [**DF19**], the authors addressed the old *question* asked by Boshernitzan. Let (X,T) be a minimal interval exchange transformation (IET) defined by a permutation of d subintervals. Due to Katok [**Kat73**] and Veech [**Vee78**], it is known that

$$|E(X,T)| \le \left[\frac{d}{2}\right].$$

Can the bound $\frac{d}{2}$ for the IET $|E(X,T)|$ be shown combinatorially using a symbolic realization (Y,S) of (X,T)?

Following [**DF19**], let us make the following assumption on the language $\mathcal{L}_X$ of a minimal subshift (X,S). A word $w \in \mathcal{L}_X$ is left special if there are distinct letters $a, a' \in \mathcal{A}$ such that aw and $a'w$ belong to $\mathcal{L}_X$. Likewise, w is right special if wb and wb' exist in the language $\mathcal{L}_X$ for distinct letters b, b'. A word w is bispecial if it is both left and right special. A bispecial word is called regular bispecial if only one left extension of w is right special and only one right extension of w is left special. The language $\mathcal{L}_X$ (or equivalently(X,S)) satisfies the regular bispecial condition if all large enough bispecial words are regular. The regular bispecial condition implies the constant growth condition above for some K. All subshifts that arise from interval exchanges satisfy this property, see [**FZ08**].

The following main result from [**DF19**] is motivated by Boshernitzan's question.

THEOREM 5.4 ([**DF19**]). *Let (X, S) be a transitive subshift satisfying the regular bispecial condition with growth constant K. Then*

$$|E(X,S)| \leq \frac{K+1}{2}.$$

We finish this subsection by pointing out an interesting application of complexity functions. It turns out that by means of the complexity function one can also estimate the number of generic measures (which are not necessarily ergodic ones). We follow here the paper [**CK19**]. We remark that the considered subshifts are not assumed to be minimal.

THEOREM 5.5 ([**CK19**]). *(1) Let (X, S) be a subshift such that*

$$\liminf_{n\to\infty} \frac{p_X(n)}{n} < K$$

for some integer K. Then (X, S) has at most $K - 1$ distinct, non-atomic, generic measures.
(2) Suppose (X, S) is a subshift satisfying the condition

$$\limsup_{n\to\infty} \frac{p_X(n)}{n} < K$$

for some integer K. If (X, S) has a generic measure μ and a generic point x_μ for which the orbit closure $(\overline{Orb_S(x_\mu)}, S)$ is not uniquely ergodic, then (X, S) has at most $K - 2$ distinct, non-atomic, generic measures.

5.2. Stationary Bratteli diagrams. In this subsection, we give an explicit description of all ergodic probability invariant measures on stationary Bratteli diagrams. Note that the class of minimal homeomorphisms which can be represented by stationary Bratteli diagrams is constituted by minimal substitution dynamical systems and odometers [**For97**, **DHS99**]. In [**BKM09**], the analogue of the above mentioned result was proved for aperiodic homeomorphisms.

The paper [**BKMS10**] contains an explicit description of all ergodic invariant probability measures on a stationary Bratteli diagram B. Let $\widetilde{F} = (\widetilde{f}_{vw})_{v,w\in V}$ be the $K \times K$ incidence matrix of the diagram B. Identify the set of vertices V_n on each level $n \geq 1$ with $\{1, \ldots, K\}$. In this subsection, by $\overline{x}$ we denote a vector, either column or row one, it will be either mentioned explicitly, or understood from the context.

The incidence matrix $\widetilde{F}$ defines a directed graph $G(\widetilde{F})$: the set of the vertices of $G(\widetilde{F})$ is equal to $\{1, \ldots, K\}$ and there is a directed edge from a vertex v to a vertex w if and only if $\widetilde{f}_{vw} > 0$. The vertices v and w are equivalent (we write $v \sim w$) if either $v = w$ or there is a path in $G(\widetilde{F})$ from v to w and also a path from w to v. Let $\mathcal{E}_1, \ldots, \mathcal{E}_m$ denote all equivalence classes in $G(\widetilde{F})$. We will also identify $\mathcal{E}_\alpha$ with the corresponding subsets of V. We write $\mathcal{E}_\alpha \succeq \mathcal{E}_\beta$ if either $\mathcal{E}_\alpha = \mathcal{E}_\beta$ or there is a path in $G(\widetilde{F})$ from a vertex of $\mathcal{E}_\alpha$ to a vertex of $\mathcal{E}_\beta$. We write $\mathcal{E}_\alpha \succ \mathcal{E}_\beta$ if $\mathcal{E}_\alpha \succeq \mathcal{E}_\beta$ and $\mathcal{E}_\alpha \neq \mathcal{E}_\beta$. Every class $\mathcal{E}_\alpha$, $\alpha = 1, \ldots, m$, defines an irreducible submatrix $\widetilde{F}_\alpha$ of $\widetilde{F}$ obtained by restricting $\widetilde{F}$ to the set of vertices from $\mathcal{E}_\alpha$. Let ρ_α be the spectral radius of $\widetilde{F}_\alpha$, i.e.

$$\rho_\alpha = \max\{|\lambda| : \lambda \in \mathsf{spec}(\widetilde{F}_\alpha)\},$$

where by $\mathsf{spec}(\widetilde{F}_\alpha)$ we mean the set of all complex numbers λ such that there exists a non-zero vector $\overline{x} = (x_v)_{v\in\mathcal{E}_\alpha}$ satisfying $\overline{x}\widetilde{F}_\alpha = \lambda\overline{x}$.

A class $\mathcal{E}_\alpha$ is called *distinguished* if

$$\rho_\alpha > \rho_\beta \quad \text{whenever} \quad \mathcal{E}_\alpha \succ \mathcal{E}_\beta \tag{5.2}$$

(in [**BKMS10**] the notion of being distinguished is defined in an opposite way because it is based on the matrix transpose to the incidence matrix).

The real number λ is called a *distinguished eigenvalue* if there exists a non-negative left-eigenvector $\overline{x} = (x_v) \in \mathbb{R}^K$ such that $\overline{x}\widetilde{F} = \lambda\overline{x}$. It is known (Frobenius theorem) that λ is a distinguished eigenvalue if and only if $\lambda = \rho_\alpha$ for some distinguished class $\mathcal{E}_\alpha$. Moreover, there is a unique (up to scaling) non-negative eigenvector $\overline{x}(\alpha) = (x_v)_{v\in V}$, $\overline{x}(\alpha)\widetilde{F} = \rho_\alpha\overline{x}(\alpha)$ such that $x_v > 0$ if and only if there is a path from a vertex of $\mathcal{E}_\alpha$ to the vertex v. The distinguished class α defines a measure μ_α on $B = (V, E)$ as follows:

$$\mu_\alpha(X_v^{(n)}) = \frac{x_v}{\rho_\alpha^{n-1}} h_v^{(n)}, \quad v \in V_n = V.$$

Theorem 5.6 ([**BKMS10**]). *Let B and $\{\mu_\alpha\}$ be as above, where α runs over all distinguished vertex classes. Then the measures $\{\mu_\alpha\}$ are exactly all ergodic $\mathcal{R}$-invariant probability measures for the stationary Bratteli diagram B.*

For instance, in Example 4.16, there is only one distinguished class of vertices which corresponds to the first vertex of the diagram on each level.

Remark 5.7. In [**BKMS10**] it was shown that non-distinguished vertex classes correspond exactly to infinite ergodic invariant measures which are finite on at least one open set.

5.3. Finite rank Bratteli diagrams. In this subsection, we give the necessary and sufficient conditions to determine the exact number of ergodic invariant probability measures on Bratteli diagrams of finite rank and describe the supports of these measures.

Definition 5.8. A Cantor dynamical system (X, S) has *topological rank* $K > 0$ if it admits a Bratteli–Vershik model (X_B, φ_B) such that the number of vertices of the diagram B at each level $V_n, n \geq 1$ is not greater than K and K is the least possible number of vertices for any Bratteli–Vershik realization.

If a system (X, S) has rank K, then, by an appropriate telescoping, we can assume that the diagram B has exactly K vertices at each level.

In [**BKMS13**], the structure of invariant measures on finite rank Bratteli diagrams is considered. In particular, it is shown that every ergodic invariant measure (finite or "regular" infinite) can be obtained as an extension from a simple uniquely ergodic vertex subdiagram. Everywhere below the term "measure" stands for an $\mathcal{R}$-invariant measure. By an infinite measure we mean any σ-finite non-atomic measure which is finite (non-zero) on some clopen set. The support of each ergodic measure for a Bratteli diagram of finite rank turns out to be the set of all paths that stabilize in some subdiagram, which geometrically can be seen as a "vertical" subdiagram, i.e. the paths will eventually stay in the subdiagram. Furthermore, these subdiagrams are pairwise disjoint for different ergodic measures. It is shown in [**BKMS13**], that for any finite rank diagram B one can find finitely many vertex subdiagrams B_α such that each finite ergodic measure on X_{B_α} extends to a (finite or infinite) ergodic measure on X_B. It is also proved that each ergodic measure (both finite and infinite) on X_B is obtained as an extension of a finite ergodic measure from some X_{B_α}. The following theorem holds.

THEOREM 5.9 ([**BKMS13**]). *Let B be a Bratteli diagram of finite rank K. The diagram B can be telescoped in such a way that for every ergodic probability measure μ there exists a subset W_μ of vertices from $\{1, \dots, K\}$ such that the support of μ consists of all infinite paths that eventually go along the vertices of W_μ only. Furthermore,*

(i) $W_\mu \cap W_\nu = \emptyset$ for different ergodic measures μ and ν;

(ii) given an ergodic probability measure μ, there exists a constant $\delta > 0$ such that for any $w \in W_\mu$ and any level n

$$\mu(X_w^{(n)}) \geq \delta;$$

(iii) the subdiagram generated by W_μ is simple and uniquely ergodic. The only ergodic measure on the path space of the subdiagram is the restriction of measure μ;

(iv) if an ergodic probability measure μ is the extension of a measure from the vertical subdiagram determined by a proper subset $W \subset \{1, \dots, K\}$, then

$$\lim_{n\to\infty} \mu(X_w^{(n)}) = 0 \text{ for all } w \notin W.$$

Condition (ii) can be used in practice to determine the support of an ergodic measure μ.

The following theorem plays an important role in the study of ergodic measures and their supports. For a finite rank Bratteli diagram B, it describes how extreme points of $\Delta_\infty^{(1)}$ determine subdiagrams of B.

THEOREM 5.10 ([**BKK**]). *Let B be a Bratteli diagram of rank K, and let B have l probability ergodic invariant measures, $1 \leq l \leq K$. Let $\{\overline{y}_1, ..., \overline{y}_l\}$ denote the extreme vectors in $\Delta_\infty^{(1)}$. Then, after telescoping and renumbering vertices, there exist exactly l disjoint subdiagrams B_i (they share no vertices other than the root) with the corresponding sets of vertices $\{V_{n,i}\}_{n=0}^\infty$ such that*

(a) for every $i = 1, ..., l$ and any $n, m > 0$, $|V_{n,i}| = |V_{m,i}| > 0$, while the set $V_{n,0} = V_n \setminus \bigsqcup_{i=1}^{l} V_{n,i}$ may be, in particular, empty;

(b) for any $i = 1, \dots, l$ and any choice of $v_n \in V_{n,i}$, the extreme vectors $\overline{y}^{(n)}(v_n) \in \Delta_n^{(1)}$ converge to the extreme vector $\overline{y}_i \in \Delta_\infty^{(1)}$.

In general, the diagram B can have up to $K - l$ disjoint subdiagrams B'_j with vertices $\{V'_{n,j}\}_{n=0}^\infty$ such that they are also disjoint with subdiagrams B_i and for any $w_n \in V'_{n,j}$, the extreme vectors $\overline{y}^{(n)}(w_n) \in \Delta_n^{(1)}$ converge to a non-extreme vector $\overline{z} \in \Delta_\infty^{(1)}$.

For a finite rank Bratteli diagram, one can describe subdiagrams that support ergodic measures in terms of the stochastic incidence matrices of the diagram. For the next theorem, we will need the following definition and notation.

DEFINITION 5.11. For a Bratteli diagram B, we say that a sequence of proper subsets $U_n \subset V_n$ defines *blocks of vanishing weights (or vanishing blocks)* in the stochastic incidence matrices F_n if

$$\sum_{w\in U_n^c, v\in U_{n+1}} f_{vw}^{(n)} \to 0, \qquad n\to\infty$$

where $U_n^c = V_n \setminus U_n$.

If additionally, for every sequence of vanishing blocks (U_n), there exists a constant $0 < C_1 < 1$ such that, for sufficiently large n,

$$\min_{v\in U_{n+1}} \sum_{u\in U_n^c} f_{vu}^{(n)} \geq C_1 \max_{v\in U_{n+1}} \sum_{u\in U_n^c} f_{vu}^{(n)}, \tag{5.3}$$

then we say that the stochastic incidence matrices F_n of B have *regularly vanishing blocks*.

Set

$$\overline{a}_j^{(n)} = \frac{1}{|V_{n,j}|} \sum_{w\in V_{n,j}} \overline{y}^{(n)}(w), \quad j = 0,1,\ldots,l,$$

where the subsets $V_{n,j}$ are defined as in Theorem 5.10. Then $\overline{a}_j^{(n)} \in \Delta_{n,j}^{(1)} := \mathrm{Conv}\{\overline{y}^{(n)}(w), w\in V_{n,j}\}$, the convex hull of the set $\{\overline{y}^{(n)}(w), w \in V_{n,j}\}$. The sets $\Delta_{n,j}^{(1)}$ are subsimplices of $\Delta_n^{(1)}$, $j = 0,1,\ldots,l$. We observe that

$$\max_{\overline{a}\in\Delta_{n,0}^{(1)}} \mathrm{dist}(\overline{a}, \Delta_\infty^{(1)}) \to 0$$

as $n\to\infty$, where $\Delta_\infty^{(1)} = \bigcap_{n=1}^\infty \Delta_n^{(1)}$.

In the theorem below, we assume that stochastic incidence matrices of a Bratteli diagram have the property of *regularly vanishing blocks* and apply it to the case when $U_n = V_{n,i}$ for some $i = 1,\ldots,l$. The blocks of the matrices corresponding to the edges that connect vertices from outside of the supporting subdiagram B_i to the vertices of B_i are the blocks of vanishing weights. Note that the second part of the theorem does not require the stochastic incidence matrices of a Bratteli diagram to have the property of regularly vanishing blocks.

Theorem 5.12 ([**BKK**]). *Let B be a Bratteli diagram of rank K such that the incidence matrices F_n have the property of regularly vanishing blocks. If B has exactly l $(1 \leq l \leq K)$ ergodic invariant probability measures, then, after telescoping, the set V_n can be partitioned into subsets $\{V_{n,1},\ldots,V_{n,l},V_{n,0}\}$ such that*

(a) $V_{n,i} \neq \emptyset$ for $i = 1,\ldots,l$;

(b) $|V_{n,i}|$ does not depend on n, i.e., $|V_{n,i}| = k_i$ for $i = 0,1,\ldots,l$ and $n \geq 1$;

(c) for $j = 1,\ldots,l$,

$$\sum_{n=1}^{\infty}\left(1 - \min_{v\in V_{n+1,j}} \sum_{w\in V_{n,j}} f_{vw}^{(n)}\right) < \infty;$$

(d) for $j = 1,\ldots,l$,

$$\max_{v,v'\in V_{n+1,j}} \sum_{w\in V_n} \left| f_{vw}^{(n)} - f_{v'w}^{(n)}\right| \to 0$$

as $n\to\infty$;

(e1) for every $w \in V_{n,0}$

$$\mathrm{vol}_l S(\overline{a}_1^{(n)}, \dots, \overline{a}_l^{(n)}, \overline{y}^{(n)}(w)) \to 0$$

as $n \to \infty$, where S is a simplex with extreme points $\overline{a}_1^{(n)}, \dots, \overline{a}_l^{(n)}, \overline{y}^{(n)}(w)$, and $\mathrm{vol}_l(S)$ *stands for the volume of S;*

(e2) for every $v \in V_{n+1,0}$ and for sufficiently large n, there exists some $C > 0$ such that, for every $j = 1, \dots, l$,

$$F_v^{(n,j)} = \sum_{w \in V_{n,j}} f_{vw}^{(n)} < 1 - C.$$

Conversely, let B be a Bratteli diagram of finite rank $K \geq 2$ with nonsingular stochastic incidence matrices (F_n). Suppose that after telescoping B satisfies conditions $(a) - (e2)$. Then B has l ergodic probability invariant measures.

Remark 5.13. Condition (d) of Theorem 5.12 guarantees that the subdiagrams B_i, $i = 1, \dots, l$ corresponding to the vertices from $V_{n,i}$, are uniquely ergodic, but it does not guarantee that the subdiagrams B_i are simple. One can reduce B_i to the smallest possible simple and uniquely ergodic subdiagrams such that the obtained subdiagrams are the same as considered in Theorem 5.9. For instance, in Example 4.16, one can take $V_{n,1} = V_n$ and $V_{n,0} = \emptyset$ for all n. After reduction, we obtain that the new set $V'_{n,1}$ consists only of the first vertex on each level n, and $V'_{n,0}$ consists of the second one. Condition (c) of Theorem 5.12 yields that for every $i = 1, \dots, l$, the extension of the unique invariant measure μ_i on B_i to the measure $\widehat{\mu}_i$ on B is finite. Conditions $(e1)$ and $(e2)$ guarantee that there are no more finite ergodic invariant measures on B except for $\widehat{\mu}_1, \dots \widehat{\mu}_l$.

The following theorem gives a criterion for the existence of K ergodic invariant probability measures on a Bratteli diagram of rank K. This criterion was proved in [**ABKK17**] for the case of Bratteli diagrams with ERS property, but actually it can be reproved in terms of stochastic incidence matrices (F_n) without the ERS property requirement.

Theorem 5.14. *Let $B = (V, E)$ be a Bratteli diagram of rank $K \geq 2$; identify V_n with $\{1, ..., K\}$ for any $n \geq 1$. Let $F_n = (f_{i,j}^{(n)})$ form a sequence of stochastic incidence matrices of B. Suppose that* rank $F_n = K$ *for all n. We write*

$$z^{(n)} = \det \begin{pmatrix} f_{1,1}^{(n)} & \dots & f_{1,k}^{(n)} \\ \vdots & \ddots & \vdots \\ f_{k,1}^{(n)} & \dots & f_{k,k}^{(n)} \end{pmatrix}.$$

Then there exist exactly K ergodic invariant measures on B if and only if

$$\prod_{n=1}^{\infty} |z^{(n)}| > 0,$$

or, equivalently,

$$\sum_{n=1}^{\infty} (1 - |z^{(n)}|) < \infty.$$

5.4. Examples.

EXAMPLE 5.15 (Stationary Bratteli diagrams). This example illustrates Theorem 5.12. For stationary Bratteli diagrams (see Subsection 5.2), we relate the distinguished classes of vertices to the subsets $V_{n,j}$ mentioned in Theorem 5.12.

PROPOSITION 5.16 ([**BKK**]). *Let $B = (V, E)$ be a stationary Bratteli diagram and $V_{n,j}$, $j = 1, \dots, l$ be subsets of vertices defined in Theorem 5.12. Then the distinguished classes α (as subsets of V) coincide with the sets $V_{n,j}$, $j = 1, \dots, l$.*

The proof of the above proposition uses the representation of the incidence matrix $\widetilde{F}$ in the Frobenius normal form (similarly to the way it was done in [**BKMS10**]):

$$F = \begin{pmatrix} \widetilde{F}_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \widetilde{F}_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \widetilde{F}_s & 0 & \cdots & 0 \\ Y_{s+1,1} & Y_{s+1,2} & \cdots & Y_{s+1,s} & \widetilde{F}_{s+1} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \ddots & \vdots \\ Y_{m,1} & Y_{m,2} & \cdots & Y_{m,s} & Y_{m,s+1} & \cdots & \widetilde{F}_m \end{pmatrix},$$

where all $\{\widetilde{F}_i\}_{i=1}^m$ are irreducible square matrices, and for any $j = s+1, \dots, m$, at least one of the matrices $Y_{j,u}$ is non-zero. All classes $\{\mathcal{E}_\alpha\}_{\alpha=1}^s$ $(s \geq 1)$, are distinguished (there is no β such that $\alpha > \beta$). For every $\alpha \geq s+1$ such that $\mathcal{E}_\alpha$ is a distinguished class and for every $1 \leq \beta < \alpha$ we have either $\mathcal{E}_\beta \prec \mathcal{E}_\alpha$ and $\rho_\beta < \rho_\alpha$, or there is no relation between $\mathcal{E}_\alpha$ and $\mathcal{E}_\beta$. Then the Perron–Frobenius theorem is used to show that the sets $V_{n,j}$, $j = 1, \dots, l$, coincide with the distinguished classes $\mathcal{E}_\alpha$.

EXAMPLE 5.17. Let B be a Bratteli diagram of rank 2 with incidence matrices

$$\widetilde{F}_n = \begin{pmatrix} n^2 & 1 \\ 1 & n^2 \end{pmatrix}, \qquad n \in \mathbb{N}.$$

Since B_2 has the ERS property (see Example 3.1), the corresponding stochastic incidence matrices are:

$$F_n = \begin{pmatrix} 1 - \dfrac{1}{n^2+1} & \dfrac{1}{n^2+1} \\ \\ \dfrac{1}{n^2+1} & 1 - \dfrac{1}{n^2+1} \end{pmatrix}.$$

Since the series

$$\sum_{n=1}^{\infty} \frac{1}{n^2+1}$$

converges, by Theorem 5.12 the diagram B has two ergodic invariant probability measures. It also is easy to see that the diagram satisfies the condition of Theorem 5.14 (see also Proposition 3.1 in [**ABKK17**]).

This example can be generalized to the case of Bratteli diagrams of rank $K \geq 2$ by using Example 4.15 and choosing the appropriate values for $f_i^{(n)}$, $i = 1, \dots, n$.

6. Infinite rank Cantor dynamical systems

In this section, we give sufficient conditions for a Bratteli diagram of infinite rank to have a prescribed (finite or infinite) number of ergodic invariant probability measures. We define a class of Bratteli diagrams of infinite rank that, in some sense, generalizes the class of Bratteli diagrams of finite rank. A diagram of this class has a prescribed number of uniquely ergodic subdiagrams such that the extension of the unique invariant measure from each subdiagram to the whole diagram is finite. Moreover, there are no other finite ergodic invariant measures for the Bratteli diagram.

6.1. A class of Bratteli diagrams of infinite rank. Let us assume that (after telescoping) every level V_n of a Bratteli diagram B admits a partition

$$V_n = \bigcup_{i=0}^{l_n} V_{n,i}, \quad n = 1, 2, \ldots,$$

into disjoint subsets $V_{n,i}$ such that $V_{n,i} \neq \emptyset$, for $i = 1, \ldots, l_n$, and $l_n \geq 1$. Moreover, let

$$L_{n+1} = \{1, \ldots, l_{n+1}\} = \bigcup_{i=1}^{l_n} L_{n+1}^{(i)},$$

where $L_{n+1}^{(i)} \neq \emptyset$ and $L_{n+1}^{(i)} \cap L_{n+1}^{(j)} = \emptyset$ for $i \neq j$, $i, j = 1, \ldots, l_n$. Hence, for every $j = 1, \ldots, l_{n+1}$, there exists a unique $i = i(j) \in \{1, \ldots, l_n\}$ such that $j \in L_{n+1}^{(i)}$. Denote

$$V_{n+1}^{(i)} = \bigcup_{j \in L_{n+1}^{(i)}} V_{n+1,j}$$

for $1 \leq i \leq l_n$.

We can interpret the sets $L_n^{(i)}$, defined above, in terms of subdiagrams. For this, select a sequence $\overline{i} = (i_1, i_2, \ldots)$ such that $i_1 \in L_1$, $i_2 \in L_2^{(i_1)}$, $i_3 \in L_3^{(i_2)}, \ldots$ and define a subdiagram $B_{\overline{i}} = (\overline{V}, \overline{E})$, where

$$\overline{V} = \bigcup_{n=1}^{\infty} V_{n,i_n} \cup \{v_0\}.$$

Now we formulate conditions $(c1)$, $(d1)$, $(e1)$ which are analogues of conditions (c), (d), (e) used in Theorem 5.12:

$(c1)$

$$\sum_{n=1}^{\infty} \left(\max_{i \in L_n} \max_{v \in V_{n+1}^{(i)}} \sum_{w \notin V_{n,i}} f_{vw}^{(n)} \right) < \infty;$$

$(d1)$

$$\max_{j \in L_{n+1}} \max_{v, v' \in V_{n+1,j}} \sum_{w \in V_n} \left| f_{vw}^{(n)} - f_{v'w}^{(n)} \right| \to 0 \text{ as } n \to \infty;$$

$(e1)$ for every $v \in V_{n+1,0}$,

$(e1.1)$

$$\sum_{w \in V_n \setminus V_{n,0}} f_{vw}^{(n)} \to 1 \quad \text{as } n \to \infty;$$

$(e1.2)$ there exists $C > 0$ such that $F_{vi}^{(n)} \le 1 - C$ for every $i = 1, \dots, l$, where

$$F_{vi}^{(n)} = \sum_{w \in V_{n,i}} f_{vw}^{(n)}.$$

Notice that the new condition (e1.1) in the case of infinite rank Bratteli diagrams is stronger than the corresponding condition (e1) in Theorem 5.12.

Let

$$\mathcal{L} = \{\bar{i} = (i_1, i_2, \dots) \ : \ i_1 \in L_1, i_{n+1} \in L_{n+1}^{(i_n)}, n = 1, 2, \dots\}.$$

We call such a sequence $\bar{i} \in \mathcal{L}$ a *chain*. We remark that a Bratteli diagram $B = (V, E)$ of finite rank has the form described in this section. We also notice that the following theorem does not require the stochastic incidence matrices of a Bratteli diagram to have the property of regularly vanishing blocks.

THEOREM 6.1 ([**BKK**]). *Let $B = (V, E)$ be a Bratteli diagram satisfying the conditions* $(c1), (d1), (e1)$. *Then:*

(1) for each $\bar{i} \in \mathcal{L}$, any measure $\mu_{\bar{i}}$ defined on $B_{\bar{i}}$ has a finite extension $\widehat{\mu}_{\bar{i}}$ on B,

(2) each subdiagram $B_{\bar{i}}$, $\bar{i} \in \mathcal{L}$, is uniquely ergodic,

(3) after normalization, the measures $\widehat{\mu}_{\bar{i}}$, $\bar{i} \in \mathcal{L}$, form the set of all ergodic invariant probability measures on B, in particular, $|\mathcal{E}_1(B)| = |\mathcal{L}|$.

6.2. Examples.

EXAMPLE 6.2 (Pascal–Bratteli diagram). In Subsection 6.1, we defined a class of Bratteli diagrams $B = (V, E)$ such that the set of all ergodic invariant probability measures coincides with the set $\mathcal{L}$ of all infinite chains $\bar{i}$. Each ergodic probability invariant measure $\widehat{\mu}_{\bar{i}}$ is an extension of a unique invariant measure $\mu_{\bar{i}}$ from the subdiagram $B_{\bar{i}}$, and the sets $X_{B_{\bar{i}}}$ are pairwise disjoint. It turns out, that the set of ergodic invariant measures for Pascal–Bratteli diagram has a different structure.

For the Pascal–Bratteli diagram, we have $V_n = \{0, 1, \dots, n\}$ for $n = 0, 1, \dots$, and the entries $\widetilde{f}_{ki}^{(n)}$ of the incidence matrix $\widetilde{F}_n$ are of the form

$$\widetilde{f}_{ki}^{(n)} = \begin{cases} 1, \text{ if } i = k \text{ for } 0 \le k < n+1, \\ 1, \text{ if } i = k-1 \text{ for } 0 < k \le n+1, \\ 0, \text{ otherwise.} \end{cases}$$

where $k = 0, \dots, n+1$, $i = 0, \dots, n$ (see [**MP05, Ver11, Ver14, FPS17**]). Moreover,

$$h_i^{(n)} = \binom{n}{i},$$

for $i = 0, \dots, n$. The entries of the corresponding stochastic matrices are F_n:

$$f_{ki}^{(n)} = \begin{cases} \dfrac{k}{n+1}, \text{ if } i = k-1 \text{ and } 0 < k \le n+1, \\ 1 - \dfrac{k}{n+1}, \text{ if } i = k \text{ and } 0 \le k < n+1, \\ 0, \text{ otherwise.} \end{cases} \tag{6.1}$$

It is known (see e.g. [**MP05**]) that each ergodic invariant probability measure has the form μ_p, $0 < p < 1$, where

$$\mu_p\left(X_i^{(n)}\right) = \binom{n}{i} p^i (1-p)^{n-i}, \quad i = 0, \dots, n.$$

PROPOSITION 6.3 ([**BKK**]). *For the Pascal–Bratteli diagram, the set $\mathcal{L}$ of all infinite chains $\overline{i}$ is empty.*

EXAMPLE 6.4 (A class of Bratteli diagrams with countably many ergodic invariant measures). In this example, we present a class of Bratteli diagrams with countably infinite set of ergodic invariant measures. Let $V_n = \{0, 1, \dots, n\}$ for $n = 0, 1, \dots$, and let $\{a_n\}_{n=0}^{\infty}$ be a sequence of natural numbers such that

$$\sum_{n=0}^{\infty} \frac{n}{a_n + n} < \infty. \tag{6.2}$$

Consider the Bratteli diagram B with $(n+2) \times (n+1)$ incidence matrices

$$\widetilde{F}_n = \begin{pmatrix} a_n & 1 & 1 & \dots & 1 & 1 \\ 1 & a_n & 1 & \dots & 1 & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & 1 & \dots & 1 & a_n \\ 1 & 1 & 1 & \dots & 1 & a_n \\ \end{pmatrix}.$$

Then

$$|r^{-1}(v)| = a_n + n$$

for every $v \in V_{n+1}$ and every $n = 0, 1, \dots$

The Bratteli diagram defined above admits an order generating the Bratteli–Vershik homeomorphism (see [**HPS92**], [**GPS95**], or [**BKY14**], [**BK16**]). In particular, we can use the so called *consecutive* ordering such that X_B has the unique minimal infinite path passing through the vertices $0 \in V_n$, $n \geq 0$ and the unique maximal infinite path passing through the vertices $n \in V_n$, $n \geq 0$. A Vershik map $\varphi_B \colon X_B \to X_B$ exists and it is minimal. Figure 2 below shows an example of such a Bratteli diagram. It is known that all minimal Bratteli–Vershik systems with a consecutive ordering have topological entropy zero (see e.g. [**Dur10**]), and hence the system that we describe in this subsection has zero topological entropy.

Denote by $B_i = (W^{(i)}, E^{(i)})$, $i = 0, 1, \dots, \infty$, the subdiagrams of B determined by the following sequences of vertices (taken consecutively from V_0, V_1, ...): for B_0, $W^{(0)} = (0, 0, 0, \dots)$; for B_i, $W^{(i)} = (0, 1, \dots, i-1, i, i, i \dots)$ for $i = 1, 2, \dots$, and for B_∞, $W^{(\infty)} = (0, 1, 2, \dots)$. Then each B_i is an odometer and $E^{(i)}$ is the set of all edges from B that belong to B_i. Let μ_i be the unique invariant (hence ergodic) probability measure on the odometer B_i. Then each measure μ_i can be extended to a finite invariant measure $\widehat{\mu}_i$ on the diagram B and it is supported by the set $\widehat{X}_{B_i}$ (see [**BKK**]). We use the same symbol $\widehat{\mu}_i$ to denote the normalized (probability) measure obtained from the extension of μ_i for $i = 0, 1, \dots, \infty$.

PROPOSITION 6.5 ([**BKK**]). *The measures $\widehat{\mu}_i$, $i = 0, 1, \dots, \infty$, form a set of all ergodic probability invariant measures on the Bratteli diagram $B = (V, E)$ defined above.*

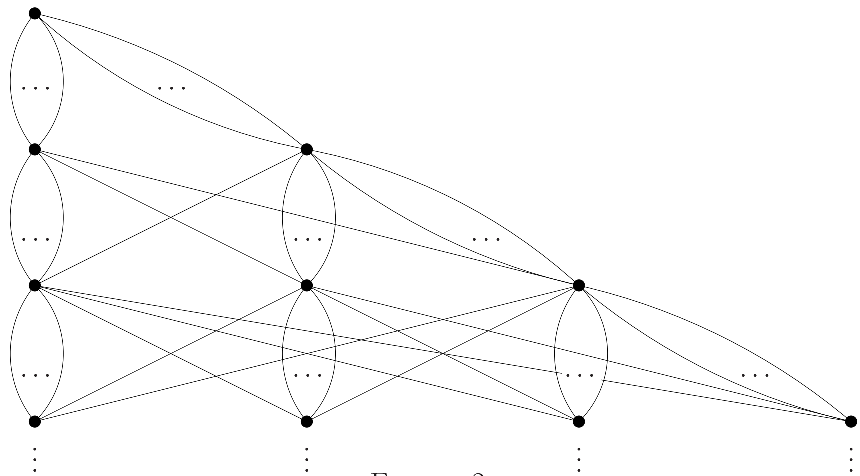


FIGURE 2

Acknowledgments. The authors are very thankful to our colleagues and collaborators for numerous fruitful and stimulating discussions. Especially, we would like to thank P. Berk, H. Bruin, T. Downarowicz, F. Durand, T. Eisner, A. Fisher, B. Frej, P. Guillon, P. Jorgensen, D. Kwietniak, J. Kwiatkowski, M. Lemańczyk, P. Muhly, P. Oprocha, K. Petersen, A. Solomko, C. Ulcigrai. We are also thankful to the referee for careful reading and useful suggestions.

References

[ABKK17] M. Adamska, S. Bezuglyi, O. Karpel, and J. Kwiatkowski, *Subdiagrams and invariant measures on Bratteli diagrams*, Ergodic Theory Dynam. Systems **37** (2017), no. 8, 2417–2452.

[BDK06] S. Bezuglyi, A. H. Dooley, and J. Kwiatkowski, *Topologies on the group of Borel automorphisms of a standard Borel space*, Topol. Methods Nonlinear Anal. **27** (2006), no. 2, 333–385. MR2237460

[BDM05] S. Bezuglyi, A. H. Dooley, and K. Medynets, *The Rokhlin lemma for homeomorphisms of a Cantor set*, Proc. Amer. Math. Soc. **133** (2005), no. 10, 2957–2964. MR2159774

[BK16] S. Bezuglyi and O. Karpel, *Bratteli diagrams: structure, measures, dynamics*, Dynamics and numbers, 2016, pp. 1–36. MR3546661

[BKK] S. Bezuglyi, O. Karpel, and J. Kwiatkowski, *Exact number of ergodic invariant measures for bratteli diagrams*, arXiv:1709.00055.

[BKK15] S. Bezuglyi, O. Karpel, and J. Kwiatkowski, *Subdiagrams of Bratteli diagrams supporting finite invariant measures*, Zh. Mat. Fiz. Anal. Geom. **11** (2015), no. 1, 3–17, 100,103. MR3364163

[BKM09] S. Bezuglyi, J. Kwiatkowski, and K. Medynets, *Aperiodic substitution systems and their Bratteli diagrams*, Ergodic Theory Dynam. Systems **29** (2009), no. 1, 37–72. MR2470626

[BKMS10] S. Bezuglyi, J. Kwiatkowski, K. Medynets, and B. Solomyak, *Invariant measures on stationary Bratteli diagrams*, Ergodic Theory Dynam. Systems **30** (2010), no. 4, 973–1007. MR2669408 (2012g:37019)

[BKMS13] S. Bezuglyi, J. Kwiatkowski, K. Medynets, and B. Solomyak, *Finite rank Bratteli diagrams: structure of invariant measures*, Trans. Amer. Math. Soc. **365** (2013), no. 5, 2637–2679. MR3020111

[BKY14] S. Bezuglyi, J. Kwiatkowski, and R. Yassawi, *Perfect orderings on finite rank Bratteli diagrams*, Canad. J. Math. **66** (2014), no. 1, 57–101. MR3150704

[BY17] Sergey Bezuglyi and Reem Yassawi, *Orders that yield homeomorphisms on Bratteli diagrams*, Dyn. Syst. **32** (2017), no. 2, 249–282. MR3638433

[Bos84] Michael Boshernitzan, *A unique ergodicity of minimal symbolic flows with linear block growth*, J. Analyse Math. **44** (1984), 77–96. MR801288

[Bos85] Michael Boshernitzan, *A condition for minimal interval exchange maps to be uniquely ergodic*, Duke Math. J. **52** (1985), no. 3, 723–752. MR808101

[Bos92] Michael D. Boshernitzan, *A condition for unique ergodicity of minimal symbolic flows*, Ergodic Theory Dynam. Systems **12** (1992), no. 3, 425–428. MR1182655

[Boy83] Mike Boyle, *Lower entropy factors of sofic systems*, Ergodic Theory Dynam. Systems **3** (1983), no. 4, 541–557. MR753922

[Bra72] O. Bratteli, *Inductive limits of finite dimensional C^*-algebras*, Trans. Amer. Math. Soc. **171** (1972), 195–234. MR0312282 (47 #844)

[Cas96] Julien Cassaigne, *Special factors of sequences with linear subword complexity*, Developments in language theory, II (Magdeburg, 1995), 1996, pp. 25–34. MR1466182

[CM15] Jon Chaika and Howard Masur, *There exists an interval exchange with a non-ergodic generic measure*, J. Mod. Dyn. **9** (2015), 289–304. MR3412151

[CFS82] I. P. Cornfeld, S. V. Fomin, and Ya. G. Sinaĭ, *Ergodic theory*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 245, Springer-Verlag, New York, 1982. Translated from the Russian by A. B. Sosinskiĭ. MR832433

[CH73] Ethan M. Coven and G. A. Hedlund, *Sequences with minimal block growth*, Math. Systems Theory **7** (1973), 138–153. MR0322838

[CK19] Van Cyr and Bryna Kra, *Counting generic measures for a subshift of linear growth*, J. Eur. Math. Soc. (JEMS) **21** (2019), no. 2, 355–380. MR3896204

[DF17] Michael Damron and Jon Fickenscher, *On the number of ergodic measures for minimal shifts with eventually constant complexity growth*, Ergodic Theory Dynam. Systems **37** (2017), no. 7, 2099–2130. MR3693122

[DF19] Michael Damron and Jon Fickenscher, *The number of ergodic measures for transitive subshifts under the regular bispecial condition*, Preprint (2019).

[Dow91] T. Downarowicz, *The Choquet simplex of invariant measures for minimal flows*, Israel J. Math. **74** (1991), no. 2-3, 241–256. MR1135237

[Dow06] T. Downarowicz, *Minimal models for noninvertible and not uniquely ergodic systems*, Israel J. Math. **156** (2006), 93–110. MR2282370

[Dow08] T. Downarowicz, *Faces of simplexes of invariant measures*, Israel J. Math. **165** (2008), 189–210. MR2403620

[DK18] Tomasz Downarowicz and Olena Karpel, *Dynamics in dimension zero: a survey*, Discrete Contin. Dyn. Syst. **38** (2018), no. 3, 1033–1062. MR3808986

[DK19] Tomasz Downarowicz and Olena Karpel, *Decisive Bratteli–Vershik models*, Studia Math. **247** (2019), no. 3, 251–271. MR3937447

[DHS99] F. Durand, B. Host, and C. Skau, *Substitutional dynamical systems, Bratteli diagrams and dimension groups*, Ergodic Theory Dynam. Systems **19** (1999), no. 4, 953–993. MR1709427

[Dur98] Fabien Durand, *A characterization of substitutive sequences using return words*, Discrete Math. **179** (1998), no. 1-3, 89–101. MR1489074

[Dur00] Fabien Durand, *Linearly recurrent subshifts have a finite number of non-periodic subshift factors*, Ergodic Theory Dynam. Systems **20** (2000), no. 4, 1061–1078. MR1779393

[Dur10] Fabien Durand, *Combinatorics on Bratteli diagrams and dynamical systems*, Combinatorics, automata and number theory, 2010, pp. 324–372. MR2759109 (2012h:37023)

[Fer96] Sébastien Ferenczi, *Rank and symbolic complexity*, Ergodic Theory Dynam. Systems **16** (1996), no. 4, 663–682. MR1406427

[Fer99] Sébastien Ferenczi, *Complexity of sequences and dynamical systems*, Discrete Math. **206** (1999), no. 1-3, 145–154. Combinatorics and number theory (Tiruchirappalli, 1996). MR1665394

[FFT09] Sebastien Ferenczi, Albert M. Fisher, and Marina Talet, *Minimality and unique ergodicity for adic transformations*, J. Anal. Math. **109** (2009), 1–31. MR2585390

[FM10] Sébastien Ferenczi and Thierry Monteil, *Infinite words with uniform frequencies, and invariant measures*, Combinatorics, automata and number theory, 2010, pp. 373–409. MR2759110

[FZ08] Sébastien Ferenczi and Luca Q. Zamboni, *Languages of k-interval exchange transformations*, Bull. Lond. Math. Soc. **40** (2008), no. 4, 705–714. MR2441143

[Fis09] Albert M. Fisher, *Nonstationary mixing and the unique ergodicity of adic transformations*, Stoch. Dyn. **9** (2009), no. 3, 335–391. MR2566907

[Fog02] N. Pytheas Fogg, *Substitutions in dynamics, arithmetics and combinatorics*, Lecture Notes in Mathematics, vol. 1794, Springer-Verlag, Berlin, 2002. Edited by V. Berthé, S. Ferenczi, C. Mauduit and A. Siegel. MR1970385

[For97] A. H. Forrest, *K-groups associated with substitution minimal systems*, Israel J. Math. **98** (1997), 101–139. MR1459849

[FPS17] Sarah Frick, Karl Petersen, and Sandy Shields, *Dynamical properties of some adic systems with arbitrary orderings*, Ergodic Theory Dynam. Systems **37** (2017), no. 7, 2131–2162.

[FO13] Sarah Bailey Frick and Nicholas Ormes, *Dimension groups for polynomial odometers*, Acta Appl. Math. **126** (2013), 165–186. MR3077947

[FP08] Sarah Bailey Frick and Karl Petersen, *Random permutations and unique fully supported ergodicity for the Euler adic transformation*, Ann. Inst. Henri Poincaré Probab. Stat. **44** (2008), no. 5, 876–885. MR2453848

[GMPS10] Thierry Giordano, Hiroki Matui, Ian F. Putnam, and Christian F. Skau, *Orbit equivalence for Cantor minimal $\mathbb{Z}^d$-systems*, Invent. Math. **179** (2010), no. 1, 119–158. MR2563761

[GPS04] Thierry Giordano, Ian Putnam, and Christian Skau, *Affable equivalence relations and orbit structure of Cantor dynamical systems*, Ergodic Theory Dynam. Systems **24** (2004), no. 2, 441–475. MR2054051

[GPS95] Thierry Giordano, Ian F. Putnam, and Christian F. Skau, *Topological orbit equivalence and C^*-crossed products*, J. Reine Angew. Math. **469** (1995), 51–111. MR1363826

[GPS99] Thierry Giordano, Ian F. Putnam, and Christian F. Skau, *Full groups of Cantor minimal systems*, Israel J. Math. **111** (1999), 285–320. MR1710743

[GJ00] Richard Gjerde and Ørjan Johansen, *Bratteli-Vershik models for Cantor minimal systems: applications to Toeplitz flows*, Ergodic Theory Dynam. Systems **20** (2000), no. 6, 1687–1710. MR1804953

[Gla03] Eli Glasner, *Ergodic theory via joinings*, Mathematical Surveys and Monographs, vol. 101, American Mathematical Society, Providence, RI, 2003. MR1958753

[Har02] Darald J. Hartfiel, *Nonhomogeneous matrix products*, World Scientific Publishing Co., Inc., River Edge, NJ, 2002. MR1878339

[HPS92] Richard H. Herman, Ian F. Putnam, and Christian F. Skau, *Ordered Bratteli diagrams, dimension groups and topological dynamics*, Internat. J. Math. **3** (1992), no. 6, 827–864. MR1194074

[JQY17] J. Janssen, A. Quas, and R. Yassawi, *Bratteli diagrams where random orders are imperfect*, Proc. Amer. Math. Soc. **145** (2017), no. 2, 721–735. MR3577873

[Kar12a] O. Karpel, *Infinite measures on Cantor spaces*, J. Difference Equ. Appl. **18** (2012), no. 4, 703–720. MR2905292

[Kar12b] O. M. Karpel, *Good measures on locally compact Cantor sets*, Zh. Mat. Fiz. Anal. Geom. **8** (2012), no. 3, 260–279, 297, 300. MR3053206

[Kat73] A. B. Katok, *Invariant measures of flows on orientable surfaces*, Dokl. Akad. Nauk SSSR **211** (1973), 775–778. MR0331438

[Kit98] Bruce P. Kitchens, *Symbolic dynamics*, Universitext, Springer-Verlag, Berlin, 1998. One-sided, two-sided and countable state Markov shifts. MR1484730

[Køu03] Petr Køurka, *Topological and symbolic dynamics*, Cours Spécialisés [Specialized Courses], vol. 11, Société Mathématique de France, Paris, 2003. MR2041676

[LM95] Douglas Lind and Brian Marcus, *An introduction to symbolic dynamics and coding*, Cambridge University Press, Cambridge, 1995. MR1369092

[Med06] Konstantin Medynets, *Cantor aperiodic systems and Bratteli diagrams*, C. R. Math. Acad. Sci. Paris **342** (2006), no. 1, 43–46. MR2193394

[MP05] Xavier Méla and Karl Petersen, *Dynamical properties of the Pascal adic transformation*, Ergodic Theory Dynam. Systems **25** (2005), no. 1, 227–256. MR2122921

[Mon09] Thierry Monteil, *Finite blocking property versus pure periodicity*, Ergodic Theory Dynam. Systems **29** (2009), no. 3, 983–996. MR2505325

[Pan84] Jean-Jacques Pansiot, *Complexité des facteurs des mots infinis engendrés par morphismes itérés*, Automata, languages and programming (Antwerp, 1984), 1984, pp. 380–389. MR784265

[Pet89] Karl Petersen, *Ergodic theory*, Cambridge Studies in Advanced Mathematics, vol. 2, Cambridge University Press, Cambridge, 1989. Corrected reprint of the 1983 original. MR1073173

[Pet12] Karl Petersen, *An adic dynamical system related to the Delannoy numbers*, Ergodic Theory Dynam. Systems **32** (2012), no. 2, 809–823. MR2901372

[PV10] Karl Petersen and Alexander Varchenko, *The Euler adic dynamical system and path counts in the Euler graph*, Tokyo J. Math. **33** (2010), no. 2, 327–340. MR2779260

[Phe01] Robert R. Phelps, *Lectures on Choquet's theorem*, Second, Lecture Notes in Mathematics, vol. 1757, Springer-Verlag, Berlin, 2001. MR1835574

[Put10] Ian F. Putnam, *Orbit equivalence of Cantor minimal systems: a survey and a new proof*, Expo. Math. **28** (2010), no. 2, 101–131. MR2671112

[Put18] Ian F. Putnam, *Cantor minimal systems*, University Lecture Series, vol. 70, American Mathematical Society, Providence, RI, 2018. MR3791491

[Que10] Martine Queffélec, *Substitution dynamical systems—spectral analysis*, Second, Lecture Notes in Mathematics, vol. 1294, Springer-Verlag, Berlin, 2010. MR2590264

[Shi18] Takashi Shimomura, *Bratteli–vershik models and graph covering models*, arXiv:1603.03940 (2018).

[Ska00] Christian Skau, *Ordered K-theory and minimal symbolic dynamical systems*, Colloq. Math. **84/85** (2000), no. part 1, 203–227. Dedicated to the memory of Anzelm Iwanik. MR1778851

[Vee78] William A. Veech, *Interval exchange transformations*, J. Analyse Math. **33** (1978), 222–272. MR516048

[Ver81] A. M. Vershik, *Uniform algebraic approximation of shift and multiplication operators*, Dokl. Akad. Nauk SSSR **259** (1981), no. 3, 526–529. MR625756

[Ver82] A. M. Vershik, *A theorem on Markov periodic approximation in ergodic theory*, Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI) **115** (1982), 72–82, 306. Boundary value problems of mathematical physics and related questions in the theory of functions, 14. MR660072

[Ver11] A. M. Vershik, *The Pascal automorphism has a continuous spectrum*, Funktsional. Anal. i Prilozhen. **45** (2011), no. 3, 16–33. MR2883236

[Ver14] A. M. Vershik, *Intrinsic metric on graded graphs, standardness, and invariant measures*, Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) **421** (2014), no. Teoriya Predstavleniĭ, Dinamicheskie Sistemy, Kombinatornye Metody. XXIII, 58–67. MR3479463

[Wal82] Peter Walters, *An introduction to ergodic theory*, Graduate Texts in Mathematics, vol. 79, Springer-Verlag, New York-Berlin, 1982. MR648108

Department of Mathematics, University of Iowa, Iowa City, 52242 IA, USA
Email address: bezuglyi@gmail.com

B. Verkin Institute for Low Temperature Physics and Engineering of the National Academy of Sciences of Ukraine, Kharkiv 61103, Ukraine
Current address: Faculty of Applied Mathematics, AGH University of Science and Technology, al. Mickiewicza 30, 30-059 Krakow, Poland
Email address: helen.karpel@gmail.com

Contemporary Mathematics
Volume **744**, 2020
https://doi.org/10.1090/conm/744/14989

Periods of abelian differentials and dynamics

Michael Kapovich

To the memory of Sergei Kolyada.

Abstract. Given a closed oriented surface S of genus ≥ 3 we describe those cohomology classes $\chi \in H^1(S, \mathbb{C})$ which appear as the period characters of abelian differentials for some choice of complex structure $\tau = \tau(\chi)$ on S consistent with the orientation. In other words, we describe the union

$$\bigcup_{\tau \in T(S)} H^{1,0}(S_\tau, \mathbb{C}),$$

where $T(S)$ is the Teichmüller space of S. The proof is based upon Ratner's solution of Raghunathan's conjecture.

1. Introduction

This paper is a slightly revised version of my preprint written in 2000 at the Max Plank Institute for Mathematics in Bonn. A few years after writing the preprint, I discovered a paper by Otto Haupt [**Hau20**], where the main result of my paper, Theorem 1.2 (including the genus 2 case), was proven by elementary methods. Another proof is contained in the preprint of Bogomolov, Soloviev and Yotov, [**BSY09**] (who also study periods of pairs and even triples of abelian differentials). In view of Haupt's paper, the main point of my work is to establish a connection of the periods of abelian differentials to ergodic theory. This connection and some of the methods used in this work were exploited by Calsamiglia, Deroin and Francaviglia in [**CDF15**] to further analyze the period map and to prove the connectivity of its fibers. In their paper they also found a mistake in my preprint, in the analysis of the genus 2 case, and gave a precise description of orbit closures in this setting. Therefore, I have removed the genus 2 case from the present paper; otherwise, it remains essentially unchanged.

Let S be a closed (i.e. compact with empty boundary) connected oriented surface of genus n. Recall that each complex structure τ on S (consistent with the orientation) determines the linear subspace $H^{1,0}(S_\tau, \mathbb{C}) \subset H^1(S, \mathbb{C})$ of complex dimension n (i.e. half of the dimension of the cohomology group). In down-to-earth terms, the subspace $H^{1,0}(S, \mathbb{C})$ consists of the period characters of abelian

2010 *Mathematics Subject Classification.* Primary 14K20, 22D40 ; Secondary 22E40, 57M50.
Key words and phrases. Abelian differentials, ergodic theory.
The author was supported in part by the National Science Foundation.

differentials $\alpha \in \Omega(S)$:

$$\chi_\alpha = \chi \in H^1(S,\mathbb{C}), \quad \chi(c) = \int_c \alpha, \quad c \in H_1(S,\mathbb{Z}).$$

In this paper we describe the subset

$$\bigcup_{\tau\in T(S)} H^{1,0}(S_\tau,\mathbb{C}) \subset H^1(S,\mathbb{C}),$$

where $T(S)$ is the Teichmüller space of S. In other words, we give a necessary and sufficient condition for a character $\chi \in H^1(S,\mathbb{C})$ to appear as the period of some abelian differential α on S_τ for *some* choice of the complex structure τ on S.

REMARK 1.1. We note the difference between this question and the Schottky problem which asks for a description of the subvariety in the Grassmannian $G(n,2n)$ whose elements are subspaces $H^{1,0}(S_\tau,\mathbb{C})$, with $\tau \in T(S)$.

Since the solution is obvious in the case $\chi = 0$ we will consider only the nontrivial characters χ. It turns out that there are precisely two topological obstructions for such χ to be the character of an abelian differential, the first is classical and is a part of the Riemann bilinear relations (see for instance [**Nar92**]); the second is less known.

To describe the first obstruction (which applies for all $n \geq 1$) recall that the Poincaré duality defines a symplectic pairing $\omega : H^1(S,\mathbb{R})^{\otimes 2} \to \mathbb{R}$. This yields a quadratic form $H^1(S,\mathbb{C}) \to \mathbb{R}$ again denoted ω:

$$\omega(\chi) := \omega(Re\chi, Im\chi).$$

If $x_1, y_1, ..., x_n, y_n$ denote the standard (symplectic) basis of $H^1(S,\mathbb{Z})$ then $\omega(\chi)$ equals

$$\sum_{i=j}^{n} Im(\overline{\chi(x_j)}\chi(y_j)).$$

The number $\omega(\chi)$ can be also described as

$$\int_S f^*(dA),$$

where dA is the area form $\frac{i}{2}dz \wedge \overline{dz}$ on $\mathbb{C}$, $f : S \to E$ is a section of the complex line bundle E over S associated with χ. (The form dA is induced on E via the projection $\tilde{S} \times \mathbb{C} \to \mathbb{C}$, where $\tilde{S}$ is the universal covering of S.)

Note that in the case when $\chi \neq 0$ is the period character of an abelian differential $\alpha \in \Omega(S)$ we have:

$$\omega(\chi) = \int_S \frac{i}{2}\alpha \wedge \bar{\alpha}$$

is the area of the surface S with respect to the singular Euclidean metric on S induced by α. Since this area has to be positive we get

Obstruction 1. If $\chi \in H^{1,0}(S_\tau)$ for some $\tau \in T(S)$ then $\omega(\chi) > 0$.

The second obstruction applies only to special characters χ and surfaces of genus $n \geq 2$. In what follows we will regard elements of $H^1(S,\mathbb{C})$ as additive characters χ on $H^1(S,\mathbb{Z})$, this way we have the *image* of χ, which is a subgroup A_χ of $\mathbb{C}$.

Obstruction 2. Suppose that the image Image(χ) of the character $\chi \in H^1(S,\mathbb{C})$ is a discrete subgroup A_χ of $\mathbb{C}$ isomorphic to $\mathbb{Z}^2$ and $n \geq 2$. Thus χ gives rise to a homomorphism

$$\chi : H^1(S,\mathbb{Z}) \to H^1(T^2,\mathbb{Z})$$

where $T^2 = \mathbb{C}/A_\chi$ is the 2-torus. This map is realized by a unique (up to homotopy) map $f : S \to T^2$. Then, for each $\chi \in H^{1,0}(S_\tau)$ the degree of f has to be at least 2.

The reason for this obstruction is that if χ is the period of some $\alpha \in \Omega(S_\tau)$ then the multivalued solution of the equation $dF = \alpha$ on the Riemann surface S_τ yields a (nonconstant) holomorphic map $f : S \to T^2$ which induces $\chi : H^1(S,\mathbb{Z}) \to H^1(T^2,\mathbb{Z})$. Since the surface S is assumed to have genus $n \geq 2$, the map f cannot be a homeomorphism, hence its degree is at least 2.

Alternatively, the second obstruction can be described as follows. Assume again that the image A_χ of the character χ is a discrete subgroup isomorphic to $\mathbb{Z}^2$. Let Area(χ) denote Area($\mathbb{C}/A_\chi$), the area of the flat torus. Then the requirement $\deg(f) \geq 2$ is equivalent to

$$\omega(\chi) \geq 2\,\mathrm{Area}(\chi).$$

We now assume that the surface S has genus $n \geq 3$. Our main result is the following:

THEOREM 1.2. *If $n \geq 3$ and $\chi \in H^1(S,\mathbb{C})$ satisfies the conditions imposed by the 1-st and the 2-nd obstruction then $\chi \in H^{1,0}(S_\tau)$ for some $\tau \in T(S)$.*

In §6 we show that if χ is a nonzero character which is not the period of any abelian differential, it is nevertheless possible to find a complex structure τ on S such that χ is the period character of a meromorphic differential with a single simple pole on S_τ. We now identify the additive group $\mathbb{C}$ with the subgroup of $PSL(2,\mathbb{C})$ consisting of translations $z \mapsto z+b, b \in \mathbb{C}$. Then we can regard χ as a representation $\rho : \pi_1(S) \to PSL(2,\mathbb{C})$. For such ρ define

$$d(\rho) := \begin{cases} 2n-2, & \text{if Obstructions 1 and 2 are satisfied,} \\ 2n, & \text{otherwise.} \end{cases} \tag{1.1}$$

We recall (see e.g. [**GKM00**]) that a *branched projective structure* σ on a complex curve S is an atlas with values in $\mathbb{S}^2$ where the local charts are nonconstant holomorphic functions (not necessarily locally univalent) and the transition maps are linear-fractional transformations (i.e. elements of $PSL(2,\mathbb{C})$). Thus near each point $z \in S$ (which we identify with $0 \in \mathbb{C}$) the local chart has the form $z \mapsto z^{m+1}$. The number $m = \deg(z)$ is called the degree of branching at z. We get the *branching divisor* D on S whose degree is called the *degree of branching* $\deg(\sigma)$. For each representation $\rho : \pi_1(S) \to PSL(2,\mathbb{C})$ there exists a complex-projective structure σ (consistent with the orientation on S) which corresponds to *some* complex structure on S, such that ρ is the holonomy of σ. We define $d(\rho)$ to be the least degree of branching for such structures. Note that for the trivial representation ρ, $d(\rho) = 2n+2$ and the branched projective structure is given by the hyperelliptic covering. In this note we compute the function $d(\rho)$ in the very special case of representations with the image in the subgroup of translations. The general case will be treated elsewhere, here we only note that in [**GKM00**] (see also [**Kap95**]) it was shown that for each representation ρ with *nonelementary image*[1], $d(\rho) \in \{0,1\}$ equals the 2-nd Stiefel–Whitney class of ρ (mod 2).

COROLLARY 1.3. *Suppose that $n \geq 3$. For each nontrivial representation $\rho : \pi_1(S) \to PSL(2,\mathbb{C})$ whose image is contained in the subgroup of translations, the function $d(\rho)$ is given by the formula (*1.1*).*

The lower bounds in this theorem are given by the Riemann–Roch theorem (see §6), while the upper bound follows from Theorems 1.2 and 6.1.

Since the map $P : \alpha \to \chi_\alpha$, which sends the abelian differential to its character, is complex-linear, it suffices to prove Theorem 1.2 for *normalized* characters, i.e. the characters χ such that $\omega(\chi) = 1$ (hence the 1-st obstruction automatically holds). We let

$$X := \{\chi \in H^1(S,\mathbb{C}) : \omega(\chi) = 1\}$$

and

$$\Sigma := X \cap \bigcup_{\tau \in T(S)} H^{1,0}(S_\tau, \mathbb{C}).$$

Let Ω denote the vector bundle over $T(S)$ whose fiber over a point $\tau \in T(S)$ consists of abelian differentials $\Omega(S_\tau)$. We let Ω' denote the submanifold in Ω consisting of abelian differentials α such that $\omega(\alpha) = 1$. We have the map

$$P : \Omega' \to \Sigma \subset X.$$

To explain the appearance of ergodic theory in the proof we will need two elementary facts about the subset Σ in X.

Fact 1. (See §2.) The map $P : \Omega' \to X$ is open. In particular, Σ is open in X.

We let $G = Sp(n) = Sp(2n,\mathbb{R})$ denote the group of linear symplectic automorphisms of the symplectic structure ω on $\mathbb{R}^{4n} = H^1(S,\mathbb{C})$. This is a simple algebraic Lie group which acts naturally on X. It is elementary that the action of G on X is transitive. The stabilizer G_χ of a point $\chi \in X$ is isomorphic to $Sp(2n-2)$. Thus $X = Sp(2n)/Sp(2n-2)$. Recall that the integer symplectic group $\Gamma = Sp(2n,\mathbb{Z})$ is a *lattice* in the group G.

Fact 2. The subset Σ is invariant under Γ.

Recall that the group of orientation-preserving diffeomorphisms $\mathrm{Diff}(S)$ acts on $H^1(S,\mathbb{C})$ through the group Γ. If $\chi \in \Sigma$ is the period character of $\alpha \in \Omega(S_\tau)$ and $\gamma \in \Gamma$ corresponds to a diffeomorphism $h : S \to S$, then $\gamma(\chi)$ is the period character of the abelian differential

$$h^*(\alpha) \in \Omega(S_{h^*(\tau)}),$$

where $h^*(\tau)$ is the pull-back of the complex structure τ via h. Thus $\gamma(\Sigma) = \Sigma$.

Combining the above two facts we see that Σ is a (nonempty) open Γ-invariant subset of X. We recall

THEOREM 1.4 (C. Moore, see [**Zim84**]). *If G is a semisimple Lie group, Γ is a lattice in G and H is a noncompact Lie subgroup in G then H acts ergodically on $\Gamma\backslash G$. Equivalently, Γ acts ergodically on G/H.*

[1]I.e. the image does not have an invariant finite nonempty subset in $\mathbb{H}^3 \cup \mathbb{S}^2$.

Thus, since $\Sigma \subset X = Sp(2n)/Sp(2n-2)$ is an open nonempty Γ-invariant subset, the complement $X - \Sigma$ has zero measure. In particular, Σ is dense in X. Ergodicity of the action $\Gamma \curvearrowright X$ implies that *generic*[2] points $\chi \in X$ have dense Γ-orbits. Our objective is to understand the *nongeneric* orbits. This is done by applying Ratner's solution of Raghunathan's conjecture. Ratner's theorem implies that there are only few types of nongeneric orbits. We will show that most of them correspond to the characters with discrete image. After we describe other orbits we will show that Obstruction 2 suffices for the existence of an abelian differential with the given period character.

Acknowledgments. During the period of this work I was partially supported by the Max–Plank Institute (Bonn) for whose hospitality I am grateful. I am also grateful to Bertrand Deroin for pointing out my mistake in the genus two case and to Curt McMullen for repeatedly asking me to publish this work. I am thankful to Pieter Moree and to the anonymous referee for numerous suggestions and corrections.

2. Geometric preliminaries

Geometric interpretation of nonzero abelian differentials α. Each nonzero abelian differential $\alpha \in \Omega(S_\tau)$ determines a singular Euclidean structure on the surface S with isolated singularities at zeroes of α, see [**Str84**]. Let $\mathrm{Zero}(\alpha) \subset S$ denote the set of zeroes of α.

The local charts for this structure are given by the branches of the indefinite integral

$$F(z) = \int_{z_0}^{z} \alpha,$$

where $z_0 \in S$ is a base-point. If α vanishes (at the order $m-1$) at a point $0 \in S$ then the local chart at 0 is a k-fold ramified covering $z \mapsto z^m$. The transition maps of the flat atlas on $S - \mathrm{Zero}(\alpha)$ are Euclidean translations. Vice-versa, suppose that we are given a flat structure on the (topological) surface S where the local charts have the form $z \mapsto z^m$, $m \geq 1$, and the transition maps away from the branch-points are Euclidean translations. This structure canonically defines a complex structure on S together with an abelian differential α obtained by the pull-back of dz via the local charts. Every such singular Euclidean structure gives rise to a *developing map* $dev : \tilde{S} \to \mathbb{C}$ where $\tilde{S}$ is the universal abelian covering of S and $H := H_1(S, \mathbb{Z})$ acts on $\tilde{S}$ by deck-transformations. The mapping dev is χ-equivariant, where $\chi : H_1(S, \mathbb{Z}) \to \mathbb{C}$ is the holonomy of the above structure (it coincides with the character of the associated abelian differential). The space $E(S)$ of the above Euclidean structures has a natural topology: the topology of uniform convergence on compacts of the developing mappings. It is easy to see that with this topology the natural bijection $E(S) \to \Omega - 0_\Omega$ is a homeomorphism. Here 0_Ω is the image of the zero-section of the bundle $\Omega \to T(S)$, i.e. 0_Ω consists of zero abelian differentials.

Matrix form of the characters. Given the standard (symplectic) basis in $H_1(S, \mathbb{Z})$, $x_1, y_1, ..., x_n, y_n$, we can identify each character $\chi : H_1(S, \mathbb{Z}) \to \mathbb{C} = \mathbb{R}^2$ with the $2 \times 2n$ matrix

$$M(\chi) := [M_1 M_2 ... M_n],$$

[2]In the measure-theoretic sense.

$$M_j = M_j(\chi) := \begin{bmatrix} a_j & b_j \\ c_j & d_j \end{bmatrix}, \quad j = 1, ..., n.$$

Here

$$\chi(x_1, ..., y_n) = (u, v)^t, \quad u = (a_1, b_1, ..., a_n, b_n), \quad v = (c_1, d_1, ..., c_n, d_n),$$

and the vectors u, v are the row-vectors of the matrix M. The group $G = Sp(2n)$ acts on the matrices M by multiplying them from the right. The matrix $M(\chi)$ is the *matrix form* of the character χ. Then we define

$$\omega_j(u, v) = det(M_j(\chi)) = \begin{vmatrix} a_j & b_j \\ c_j & d_j \end{vmatrix}, \ j = 1, ..., n;$$

it follows that $\omega(u, v) = \sum_j \omega_j(u, v)$. The group $SL(2) = Sp(2)$ acts on the characters χ by multiplying their matrices from the left. It is clear that this action commutes with the action of $Sp(2n, \mathbb{Z}) \subset G$ and that it preserves each determinant $\omega_j(\chi)$.

Lemma 2.1. *$Sp(2)\Sigma = \Sigma$.*

Proof. Suppose that $\chi \in \Sigma$ is the period character of an abelian differential corresponding to a singular Euclidean structure σ. Take $A \in Sp(2)$. Composing coordinate charts of σ with A deforms σ to a new singular Euclidean structure of the same area. The holonomy of this structure is the composition $A \circ \chi$. Hence $A\chi \in \Sigma$. □

Lemma 2.2. *Suppose that $\chi = (u, v)$ and $u, v \in \mathbb{R}^{2n}$ span a 2-dimensional rational subspace (i.e. a subspace which admits a rational basis). Then the $\mathbb{Z}$-module $\mathcal{M}$ generated by the columns of the matrix $M(\chi)$ has rank 2, i.e. is discrete as a subgroup of $\mathbb{R}^2$.*

Proof. The action of $GL(2)$ by multiplication from the left on the matrix $M(\chi)$ preserves the rank of $\mathcal{M}$. Since Span(u, v) is a rational subspace there exists a matrix $A \in GL(2)$ such that the matrix $AM(\chi)$ has integer entries. The rank of the $\mathbb{Z}$-module generated by its columns is clearly 2. □

Define

$$X_+ := \{\chi \in X : \omega_j(\chi) > 0, j = 1, ..., n\}.$$

Our strategy in dealing with the *nongeneric characters* $\chi \in X$ is to find $\gamma \in Sp(2n, \mathbb{Z})$ such that $\omega_j(\gamma\chi) > 0$, $j = 1, ..., n$, i.e. $\gamma\chi \in X_+$. As we will see in Theorem 2.3 the existence of such γ would imply that χ belongs Σ (i.e. that χ is the period character of an abelian differential).

Theorem 2.3. *$X_+ \subset \Sigma$.*

Proof. Let $(u, v) \in X_+$, $u = (a_1, b_1 ..., a_n, b_{2n}), v = (c_1, d_1, ..., c_n, d_n)$. We let $z_j := (a_j, c_j), w_j := (b_j, d_j) \in \mathbb{R}^2$, $j = 1, ..., n$. Each pair of vectors (z_j, w_j) determines a fundamental parallelogram P_j in $\mathbb{R}^2$ for the lattice generated by z_j, w_j. Using parallel translations place these parallelograms such that $P_j \cap P_{j+1}$ has nonempty interior, $j = 1, ..., n-1$. Then for each pair of parallelograms P_j, P_{j+1} ($j = 1, ..., n-1$) cut both P_j, P_{j+1} open along common segments β_j and then glue them along the resulting circles. Call the result Φ. See Figure 1.

Finally, for each parallelogram P_j identify the opposite sides via a parallel translation. The result is a surface S, equipped with the projection $\delta : \tilde{S} \to \mathbb{C}$ where $\tilde{S}$ is the universal abelian covering. The surface Φ is the fundamental domain for the

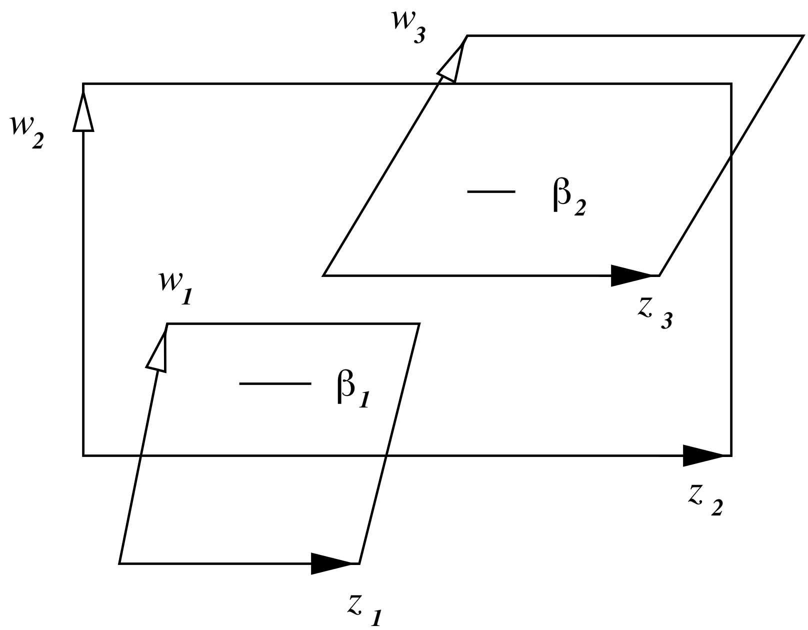


FIGURE 1

action of $H_1(S,\mathbb{Z})$ on $\tilde{S}$ via deck transformations. The restriction $\delta|\Phi : \Phi \to \mathbb{C}$ is the obvious projection. Note that δ is a local homeomorphism away from the translates of the end-points of the segments β_j. Near the end-points of such segments the mapping δ is a 2-fold ramified covering. The abelian differential α on S is obtained by taking the pull-back of dz from $\mathbb{C}$ to $\tilde{S}$ via δ and then projecting it to S. The edges of the parallelograms P_j correspond to the standard generators of $H_1(S,\mathbb{Z})$. It is clear that the periods of α over the generators of $H_1(S,\mathbb{Z})$ are given by evaluation of χ on these generators. □

The above lemma implies that it suffices to show that $\Gamma\chi \cap X_+ \neq \emptyset$ to prove that $\chi \in \Sigma$. Note however that there are characters in Σ which do not belong to the orbit ΓX_+. These are the characters with the discrete image $A_\chi \cong \mathbb{Z}^2$ such that

$$\frac{\omega(\chi)}{\mathrm{Area}(\mathbb{C}/A_\chi)} < n.$$

To find abelian differentials corresponding to such characters we need another construction that we describe below.

LEMMA 2.4. *Suppose that the character χ has the matrix form*

$$[M_1M_2...M_n], M_1 = \begin{bmatrix} a_1 = \omega(\chi) & 0 \\ 0 & 1 \end{bmatrix}, M_j = \begin{bmatrix} a_j & 0 \\ 0 & 0 \end{bmatrix}, j = 2, ..., n,$$

where $0 < a_j < a_1$, $j = 2, ..., n$. Then $\chi \in \Sigma$.

PROOF. Similarly to the previous lemma we construct a complex structure and an abelian differential by gluing certain polygons. Let P_1 be the fundamental rectangle for the group generated by the vectors z_1, w_1 which are the columns of M_1. Inside P_1 choose pairwise disjoint horizontal segments $\beta_j, \beta'_j, j = 2, .., n$, such that the translation via $[a_j 0]$ sends β_j to β'_j. We then cut P_1 open along the segments β_j, β'_j and identify the resulting circles via the translations by $[a_j 0]$, $j = 2, .., n$. Finally, glue the sides of P_1 via the horizontal translations, see Figure 2. Analogously to the previous lemma we get a singular Euclidean structure with the holonomy χ. The singular points of this structure correspond to the end-points of the segments β_j (the total angle at each of these points is 4π). □

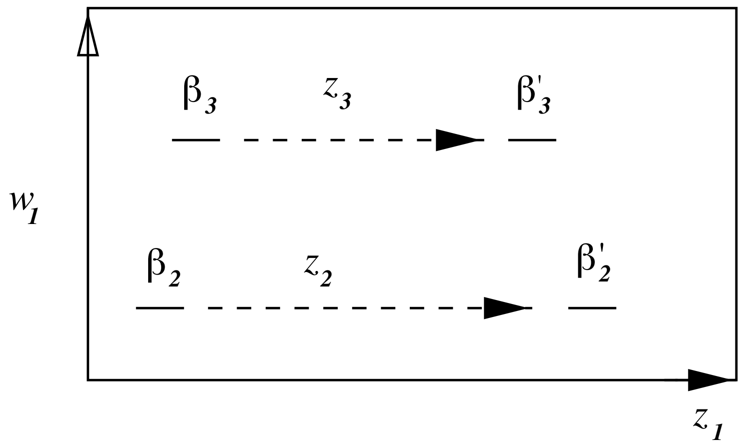


FIGURE 2

LEMMA 2.5. *Suppose that $u, v \in \mathbb{Z}^4$ are vectors such that $\omega(u, v) = 1$. Then this pair of vectors can be completed to an integer symplectic basis in $\mathbb{R}^4$.*

PROOF. Let $W := \mathrm{Span}(u, v)$. Recall that the symplectic projection $\mathrm{Proj}_W(z)$ of a vector z to W is given by

$$\mathrm{Proj}_W(z) = \omega(z, v)u - \omega(z, u)v.$$

Hence $\ker(\mathrm{Proj}_W) = W^\perp$ is a rational subspace in $\mathbb{R}^4$ and we choose a basis $p, q \in W^\perp$ such that the vectors p, q generate the abelian group $\mathbb{Z}^4 \cap W^\perp$. The vectors u, v, p, q generate the group $\mathbb{Z}^4$ since the symplectic projection of $\mathbb{Z}^4$ to W and $W^\perp$ is contained in $\mathbb{Z}^4 \cap W$ and $\mathbb{Z}^4 \cap W^\perp$ respectively. It follows that $\omega(p, q) = 1$ and x, y, p, q form an integer symplectic basis in $\mathbb{R}^4$. □

LEMMA 2.6. *Suppose that $u \in \mathbb{R}^{2n}$ is a nonzero vector. Then there exists $\gamma \in \Gamma$ such that no coordinate of $\gamma(u)$ is zero. If $u, v \in \mathbb{R}^{2n}$ are such that $\omega(u, v) > 0$, then there exists $\gamma \in \Gamma$ such that $\omega_j(\gamma(u), \gamma(v)) \neq 0$ for each $j = 1, ..., n$.*

PROOF. The projection $Sp(2n) \to \mathbb{R}^{2n} - 0$ given by $g \mapsto g(\overrightarrow{e_1})$ is a real algebraic morphism. The union

$$\bigcup_{j=1}^{n} \{x \in \mathbb{R}^{2n} : x_j = 0\}$$

is a proper (real) algebraic subvariety, hence its inverse image Y in $G = Sp(2n, \mathbb{R})$ is again a proper algebraic subvariety. Since Γ is Zariski dense in G we conclude that Y is not Γ-invariant. The proof of the second assertion is similar and is left to the reader. □

Recall that Ω denotes the vector bundle over the Teichmüller space $T(S)$ where the fiber over a point τ consists of abelian differentials on the Riemann surface S_τ; 0_ω denotes the image of the zero section of Ω. We have the period map $P : \Omega \to H^1(S, \mathbb{C})$, $\alpha \mapsto \chi_\alpha$.

The following theorem is a variation on the Hejhal–Thurston Holonomy theorem, see [**Hej75**], [**Thu81**], and [**ECG87**], [**Gol87**]. See also [**GKM00**, Section 12] for an alternative argument.

Theorem 2.7. *(The Holonomy Theorem.) The restriction mapping $P : \Omega - 0_\Omega \to H^1(S, \mathbb{C})$ is open.*

Proof. To prove this theorem we be using a geometric description of the nonzero abelian differentials α given in the beginning of this section. Let $\sigma \in E(S)$ be a singular Euclidean structure with the period character χ. Let $f : \tilde{S} \to \mathbb{C}$ denote the developing mapping of σ. Suppose that $\chi_k : H_1(S, \mathbb{Z}) \to \mathbb{C}$ is a sequence of characters converging to χ. Our goal is to find (for large k) structures $\sigma_k \in E(S)$ with the period characters χ_n and such that $\lim_k \sigma_k = \sigma$.

Choose a triangulation T of S such that each edge is a geodesic arc with respect to the singular Euclidean structure σ and each simplex is contained in a coordinate neighborhood of σ. We will assume that each singular point of σ is a vertex of this triangulation. Lift this triangulation to a triangulation $\tilde{T}$ of $\tilde{S}$ of S. Pick a finite collection $\Delta_1, ..., \Delta_M$ of 2-simplices in $\tilde{T}$, one for each H-orbit. Let $g_i, i = 1, ..., N$, be the elements of the deck-transformation group H, such that

$$g_i(\cup_j \Delta_j) \cap \cup_j \Delta_j \neq \emptyset.$$

Let C be a compact subset of $\tilde{S}$ whose interior contains both $D := \cup_j \Delta_j$ and its images under g_i's. For each χ_k we construct a continuous χ_k-equivariant mapping $f_k : D \to \mathbb{C}$ such that:

(i) f_k maps each 2-simplex homeomorphically to a Euclidean 2-simplex in $\mathbb{C}$.

(ii) f_k's converge to $f|D$ uniformly on compacts.

Finally, extend each f_k to a χ_k-equivariant mapping $f_k : \tilde{S} \to \mathbb{C}$. It remains to show that each mapping f_k is a local homeomorphism for large k (away from the singular points) and is the $m(x)$-fold ramified covering at each point where f is such a covering. It suffices to check this for points in D.

(a) If $x \in \text{int}(C)$ belongs to the interior of a 2-simplex in $\cup_i g_i D$, then the claim follows since each f_n is a homeomorphism on each simplex.

(b) Suppose x belongs to the interior of a common arc η of two 2-simplices Δ, Δ' in $\cup_i g_i D$. Since f is a local homeomorphism, $f(\Delta), f(\Delta')$ lie (locally) on different sides of the segment $f(\eta) \subset \mathbb{C}$. Therefore the same holds for f_k if k is sufficiently large. Thus, f_k does not "fold" along the arc η and is a local homeomorphism at x.

(c) Lastly, if x is a vertex of a simplex, then the degree of f at x equals $m(x)$, hence for large k, the degree of f_k at x is $m(x)$ and it follows from (b) that f_k is a $m(x)$-fold ramified covering at x.

Equivariance of f_k's implies that they converge to f uniformly on compacts. □

Line stabilizers in $Sp(2n)$. In what follows we will need a description of the subgroups B in $Sp(2n)$ with invariant line $L \subset \mathbb{R}^{2n}$. Let $V \subset \mathbb{R}^{2n}$ be a 2-dimensional symplectic subspace containing L. To describe the structure of the group B we have to recall several facts about *Heisenberg groups*. Consider the $2n-2$-dimensional symplectic vector space $(V, \omega|V)$. The Heisenberg group corresponding to this data is the $2n-1$-dimensional Lie group which fits into the short exact sequence

$$1 \to \mathbb{R} \to H_{2n-1} \to V \to 1,$$

where V is treated as the abelian (additive) Lie group. The normal subgroup $\mathbb{R}$ is central in H_{2n-1}. If $g, h \in H_{2n-1}$ project to the vectors $x, y \in V$ then $[g, h] = \omega(x, y) \in \mathbb{R}$. The *Heisenberg dilation* on this group is the action of the (multiplicative) group $\mathbb{R}_+$ on H_{2n-1} such that $t \in \mathbb{R}_+$ acts on the center $\mathbb{R} \subset H_{2n-1}$ via multiplication by t^2 and acts on V via multiplication by t. Given this one defines the Lie group $H_{2n-1} \rtimes \mathbb{R}_+$ where $\mathbb{R}_+$ acts on the Heisenberg group via Heisenberg dilation. One can show that the resulting Lie group acts simply-transitively on the complex-hyperbolic space $\mathbb{CH}^n$ of the complex dimension n, however we will not need this fact. What we will use is the following elementary lemma.

Lemma 2.8. *The $2n$-dimensional Lie group $CH_{2n} := H_{2n-1} \rtimes \mathbb{R}_+$ contains no lattices.*

Proof. Suppose that Δ is a discrete subgroup of $H_{2n-1} \rtimes \mathbb{R}_+$ with the quotient $M = H_{2n-1} \rtimes \mathbb{R}_+/\Delta$. The unit speed flow on $H_{2n-1} \rtimes \mathbb{R}_+$ along the $\mathbb{R}_+$-factor is volume-expanding and Δ-invariant. Hence it yields a volume-expanding flow on M. It follows that $\mathrm{Vol}(M) = \infty$. □

We are now ready to describe the structure of B. The group B preserves the span $L + V$ of L and V, the projection $L + V \to V$ along the L-factor transfers the action of B to the action of the symplectic group $Sp(2n-2)$ on V. The kernel of the homomorphism $B \to Sp(2n-2)$ is the group $CH_{2n} = H_{2n-1} \rtimes \mathbb{R}_+$. Here the $\mathbb{R}_+$-factor acts trivially on V and as the maximal torus in $Sp(2) \curvearrowright V^{\perp}$ preserving L. The center $\mathbb{R}$ of the Heisenberg group H_{2n-1} is the kernel of the action $B \curvearrowright L + V$. The whole group B splits as the semidirect product $CH_{2n} \rtimes Sp(2n-2)$, where $Sp(2n-2)$ acts by conjugation on the V-factor of H_{2n-1} the same way it acts on the vector space V. The subgroup $Sp(2n-2)$ commutes with the subgroup $B_0 := \mathbb{R} \rtimes \mathbb{R}_+$, where $\mathbb{R}$ is the center of H_{2n-1}. The proof of these assertions is a straightforward linear algebra computation and is left to the reader.

Definition 2.9. The group H_{2n-1} is called the *Heisenberg group* associated to the flag (V, L) in $(\mathbb{R}^{2n}, \omega)$, where V is a 2-dimensional symplectic subspace and L is a line.

3. Ratner's Theorem

Let G be a reductive algebraic Lie group and $U \subset G$ be a connected subgroup generated by unipotent elements[3]. Suppose $\Gamma \subset G$ is a lattice, i.e. a discrete subgroup with the quotient $\Gamma\backslash G$ of finite volume (with respect to the left-invariant measure on G). Important examples of lattices in algebraic Lie groups G defined over $\mathbb{Q}$ are given by the *arithmetic groups*, i.e. subgroups commensurable with $G_{\mathbb{Z}}$, the group of integer points in G. The group U acts by right multiplications on the manifold $M = \Gamma\backslash G$. On the other hand, the group Γ acts by the left multiplication on the manifold $X = G/U$. Given $g \in G$ we let $[g]$ denote its projection to M.

[3]I.e. elements whose adjoint action on the Lie algebra of G is unipotent.

THEOREM 3.1 (M. Ratner, see [**Rat91**, **Rat95**]). *Under the above conditions for each $g \in G$ the closure (in the classical topology) of $[g]U$ in M is "algebraic". More precisely, there exists a Lie subgroup $H \subset G$ such that*

- $\overline{[g]U} = [g]H$.
- $H^g \cap \Gamma$ *is a lattice in* $H^g := gHg^{-1}$.

This result is known as *Raghunathan's Conjecture*. Special cases of this conjecture were proven before Ratner by Dani [**Dan86**] and Margulis [**Mar89**]. Actually, Ratner's theorem does more than what is stated above: it describes Γ-invariant ergodic measures on M and uses the ergodic framework to prove Raghunathan's Conjecture. We note that the group H may not be connected, however if $H(0)$ is the connected component of the identity in H, then $H(0) \cap \Gamma$ is still a lattice in $H(0)$. Below we reformulate Ratner's theorem in terms of the action of Γ on G/U. Let $g \in G$ be the element which projects to x. Then

$$\overline{\Gamma g U} = \Gamma g H = \Gamma H^g g.$$

Hence

COROLLARY 3.2. *Suppose that $X := G/U$ and $x = gU \in X$. Then the closure of Γx in X equals the H^g-orbit of x in X, where H^g is a Lie subgroup of G such that $H^g \cap \Gamma$ is a lattice in H.*

Note that $gUg^{-1} = G_x$ is the stabilizer of x in G. By taking the connected component of the identity we get:

COROLLARY 3.3. *The closure $\overline{\Gamma x}$ in X contains the orbit $\Gamma F_x x$, where F_x is a connected Lie subgroup of G which contains G_x and $\Gamma \cap F_x$ is a lattice in F_x.*

Ratner's theorem gives a tool for describing the *exceptional* orbits for the Γ-action on X, still, some work has to be done by analyzing various Lie subgroups $F_x \subset G$ which might appear.

We now specialize to the case $G = Sp(2n, \mathbb{R})$, the automorphism group of the standard symplectic form ω:

$$\omega(a_1, b_1, ..., a_n, b_n) = \sum_{j=1}^{n} a_j b_{j+1} - a_{j+1} b_j,$$

and $X \subset (\mathbb{R}^{2n})^2$ consists of the pairs of vectors u, v such that $\omega(u, v) = 1$.

The stabilizer U of the point $(\overrightarrow{e_1}, \overrightarrow{e_2}) \in X$ is the group $Sp(2n-2, \mathbb{R})$ embedded in G as the subgroup of block-diagonal matrices:

$$\begin{bmatrix} 1 & 0 & 0\dots0 \\ 0 & 1 & 0\dots0 \\ 0 & 0 & Sp(2n-2) \end{bmatrix}.$$

Although the group U is not unipotent itself, it is generated by unipotent elements, hence Ratner's theorem applies. Recall that $\Gamma = Sp(2n, \mathbb{Z})$ is a lattice in G, we also note that $\Gamma \cap U$ is a lattice in U as well. In the rest of the paper we will use the notation $U' = G_\chi$ to denote the stabilizer of the point $\chi \in X$.

Connected Lie subgroups of G containing U. To apply Ratner's theorem we have to know which Lie subgroups of G contain the Lie subgroup U' (conjugate to U). We will list all *maximal* subgroups containing U. Recall that a connected Lie subgroup $G_1 \subset G$ is said to be *maximal* if it is not contained in any proper connected Lie subgroup $G_2 \subset G$. We will use a classification of maximal subgroups of classical complex Lie groups done by Dynkin [**Dyn52**] (the real case was carried out by Karpelevich [**Kar55**]). In our case the classification of maximal subgroups of $Sp(2n, \mathbb{C})$ easily implies (via the complexification) the needed result for the group of real points $Sp(2n, \mathbb{R})$.

THEOREM 3.4 (E. Dynkin, see Ch. 6, Theorems 3.1, 3.2 in [**GOV94**]). *Suppose that $H \subset Sp(2n, \mathbb{C})$ is a maximal connected Lie subgroup. Then one of the following holds:*

(a) H is a maximal parabolic subgroup of $Sp(2n, \mathbb{C})$.

(b) H is conjugate to the subgroup $Sp(k, \mathbb{C}) \times Sp(N-k, \mathbb{C})$.

(c) H is conjugate to $Sp(s, \mathbb{C}) \otimes SO(t, \mathbb{C})$ where $2n = st$, $s \geq 2, t \geq 3$, $t \neq 4$ or $s = 2, t = 4$.

Note that in our situation H contains $U \cong Sp(2n-2, \mathbb{C})$, hence we can ignore the case (c). In the case (b) the only possibility is that F is conjugate to the group $Sp(2, \mathbb{C}) \times Sp(2n-2, \mathbb{C})$. In the case (a) the group H has to preserve a complex line in $\mathbb{C}^{2n}$.

We let $\chi = (u, v)$, $u, v \in \mathbb{R}^{2n}$ are such that $\omega(u, v) = 1$. Let V denote $\mathrm{Span}(u, v)$. The group $U' = G_\chi \cong Sp(2n-2, \mathbb{R})$ fixes the vectors u, v. This group also acts as the full group of linear symplectic automorphisms of the symplectic complement $V^\perp \cong \mathbb{R}^{2n-2}$ of V. The maximal subgroups of G which contain U' are:

(1) The group $H = Sp(V) \times U'$, where $Sp(V) \cong Sp(2, \mathbb{R})$ is the group of automorphisms of V. (The semisimple case.)
(2) The maximal parabolic subgroup H of G which has an invariant line $L \subset \mathbb{R}^{2n}$. (The non-semisimple case.) We note that in this case L is necessarily contained in V.

Recall that in each case we have to find connected subgroups $F_\chi \subset H$ which contain $G_\chi = U'$ and such that $F_\chi \cap \Gamma$ is a lattice in F_χ.

4. The semisimple case

In this case the group $F_\chi \subset Sp(V) \times U'$ containing U', splits as the direct product

$$F_\chi \cong S \times Sp(2n-2),$$

where $S \subset Sp(2)$. We will need the following

THEOREM 4.1 (See e.g. [**Mar91**]). *Suppose that F_1, F_2 are simple real algebraic Lie groups such that their complexifications do not have isomorphic Lie algebras. Then any lattice $\Delta \subset F_1 \times F_2$ is reducible, i.e. $\Delta \cap F_i$ is a lattice for each $i = 1, 2$.*

We also recall (see [**Rag72**, Corollary 8.28]):

THEOREM 4.2 (M. Raghunathan, J. Wolf). *Suppose that F is a connected Lie group whose semisimple part contains no compact factors acting trivially on the radical $R(F)$ of F. Then each lattice $\Delta \subset F$ intersects the radical $R(F)$ along a sublattice in $R(F)$. Moreover, the projection of Δ to $F/R(F)$ is a lattice in this Lie group.*

In our case the group S is either solvable or equals $Sp(2)$, hence combining the two above theorems we conclude that either:

(i) $\Gamma \cap U'$ is a lattice, or

(ii) $n = 2$, $F_\chi \cong Sp(2) \times Sp(2)$ and $\Gamma \cap U'$ is not a lattice[4].

In view of the assumption that S has genus ≥ 3, we are considering here only case (i), when $\Gamma \cap U'$ is a lattice.

By the Borel density theorem (see e.g. [**Zim84**]) the intersection $U' \cap \Gamma$ is Zariski dense in U', in particular it contains a diagonalizable matrix $A \in Sp(2n)$ which has the eigenvalue 1 of the multiplicity 2. Since A has rational entries, the kernel $\ker(A-I)$ is a rational subspace. We recall that the group U' is the pointwise stabilizer of the linear subspace $\mathrm{Span}(u, v)$ of $\mathbb{R}^{2n}$ spanned by $u = Re(\chi), v = Im(\chi)$. Hence $\mathrm{Span}(u, v)$ is a rational subspace of $\mathbb{R}^{2n}$.

Lemma 2.2 thus implies that the image A_χ of the character $\chi : H_1(S, \mathbb{Z}) \to \mathbb{C}$ is a discrete subgroup of $\mathbb{C}$ isomorphic to $\mathbb{Z}^2$. Moreover, without loss of generality we can assume that A_χ is the standard integer lattice in $\mathbb{C}$ (see Section 2). This might require scaling $\omega(\chi)$ by a positive real number.

We recall that $\omega(u, v) > 0$, where $\chi = (u, v)$,

$$u = (a_1, b_1, ..., a_n, b_n), \quad v = (c_1, d_1, ..., c_n, d_n), \quad a_j, b_j, c_j, d_j \in \mathbb{Z}.$$

LEMMA 4.3. *There exists $\gamma \in \Gamma$ such that the character $\gamma\chi = \chi' = (u', v')$ satisfies:*

(i) $\omega_1(u', v') > 0$.

(ii) $\omega_j(u', v') = 0$ for each $j \geq 2$ and, moreover,

$$M_j(\chi') = \begin{bmatrix} a'_j & b'_j \\ c'_j & d'_j \end{bmatrix} = \begin{bmatrix} a'_j & 0 \\ 0 & 0 \end{bmatrix}, a'_j \geq 0.$$

PROOF. We recall that without loss of generality we can start with (u, v) such that for each $j = 1, ..., n$, $\omega_j(u, v) \neq 0$ or

$$M_j(\chi) = \begin{bmatrix} a_j & 0 \\ 0 & 0 \end{bmatrix}.$$

(Of course, in the beginning of the induction the latter case does not occur.) After multiplying (u, v) by a matrix in $\Gamma \cap Sp(2) \times ... \times Sp(2)$ we can assume that every matrix

$$M_j(\chi) = \begin{bmatrix} a_j & b_j \\ c_j & d_j \end{bmatrix} = \begin{bmatrix} a_j & 0 \\ 0 & d_j \end{bmatrix}$$

is diagonal. We now argue inductively. Suppose that $j \in \{2, .., n\}$. We let $d'_j := d_j/\gcd(|d_1|, |d_j|)$. Then there are integers α_j, β_j such that $\alpha_j d'_j - \beta_j d'_1 = 1$. It follows that

$$\begin{bmatrix} \alpha_j & 0 & \beta_j & 0 \\ a_j & d'_j & -a_1 & -d'_1 \end{bmatrix} \begin{bmatrix} a_1 & 0 \\ 0 & d_1 \\ a_j & 0 \\ 0 & d_j \end{bmatrix} = \begin{bmatrix} \alpha_j a_1 + \beta_j a_j & 0 \\ 0 & 0 \end{bmatrix}.$$

[4]We note that the group $Sp(2) \times Sp(2)$ contains irreducible lattices, namely the Hilbert modular groups.

Note that the row vectors p, q of the first matrix in the above formula are such that $\omega(p,q) = 1$. Hence, according to Lemma 2.5, there exists a matrix

$$A = \begin{bmatrix} * & * & * & * \\ * & * & * & * \\ \alpha & 0 & \beta & 0 \\ a_2 & d_2' & -a_1 & -d_1' \end{bmatrix}$$

which belongs to $Sp(4,\mathbb{Z})$. We extend the matrix A to a matrix $g \in Sp(2n,\mathbb{Z})$ which preserves all the coordinates except a_1, b_1 and a_j, b_j. Then the character $\chi' = g\chi$ has $\omega_j(\chi') = 0$. Continuing inductively we find $h \in \Gamma$ such that the character $h\chi$ satisfies:

$$\omega_j(h\chi) = 0, j = 2, 3, ..., n.$$

Note that $\omega_1(h\chi) = \omega(h\chi) = \omega(\chi) > 0$. Recall that Image$(\chi) = \mathbb{Z} \times \mathbb{Z}$. Hence $b_1' = \chi(y_1) = 1$, since all other generators $x_1, x_2, y_2, ...$ of $H_1(S,\mathbb{Z})$ are mapped by $h\chi$ to the real numbers. Finally, to obtain $\gamma\chi$ as required by the lemma, we multiply $h\chi$ by a diagonal symplectic matrix with diagonal entries in $\{\pm 1\}$ to get $a_j \geq 0$ for $j = 2, ..., n$. $\square$

We again use the notation χ for the character χ' obtained in the previous lemma.

LEMMA 4.4. *There exists $\gamma \in \Gamma$ such that that the character $\gamma\chi$ satisfies:*
(i)

$$M_1(\gamma\chi) = \begin{bmatrix} a_1 = \omega(\chi) & 0 \\ 0 & 1 \end{bmatrix}.$$

(ii) For each $j \geq 2$,

$$M_j(\gamma\chi') = \begin{bmatrix} a_j' & 0 \\ 0 & 0 \end{bmatrix}, 0 \leq a_j' < a_1.$$

PROOF. For each $j \geq 2$ there exists $t_j \in \mathbb{Z}$ such that $0 \leq a_j' := a_j - t_j a_1 < a_1$. Then form the symplectic matrix

$$\gamma = \left[\begin{array}{cc|cc|cc|c|cc} 1 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & t_2 & 0 & t_3 & \dots & 0 & t_n \\ -t_2 & 0 & 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ -t_3 & 0 & 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -t_n & 0 & 0 & 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 1 \end{array}\right].$$

The reader will note that this matrix belongs to the Heisenberg subgroup of $Sp(2n)$ associated to the flag $(\mathrm{Span}(e_1, e_2), \mathrm{Span}(e_2))$. Then $\gamma\chi$ has the requires properties:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & t_j \\ -t_j & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_1 & 0 \\ 0 & 1 \\ a_j & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} a_1 & 0 \\ 0 & 1 \\ a_j' & 0 \\ 0 & 0 \end{bmatrix}. \quad \square$$

We note that for some j we might have $a_j' = 0$. However, since $\omega(\chi) \geq 2 = \mathrm{Area}(\mathbb{C}/\mathbb{Z}^2)$ we conclude that there exists at least one $j \geq 2$ such that $a_j > 0$.

Rename this index j to make it equal to 2. Rename $\chi' = \gamma\chi$ back to χ and a'_j back to a_j, $j = 2, ..., n$.

LEMMA 4.5. *There exists $\gamma \in \Gamma$ such that that the character $\gamma\chi$ satisfies:*
(i)

$$M_1(\gamma\chi) = \begin{bmatrix} a_1 = \omega(\chi) & 0 \\ 0 & 1 \end{bmatrix}.$$

(ii) For each $j \geq 2$,

$$M_j(\gamma\chi) = \begin{bmatrix} a'_j & 0 \\ 0 & 0 \end{bmatrix}, \quad 0 < a'_j < a_1.$$

PROOF. The required matrix γ belongs to the Heisenberg group associated to the flag $(\mathrm{Span}(e_3, e_4), \mathrm{Span}(e_4))$. For each j such that $a_j \neq 0$ the multiplication by γ will not change a_j at all. Suppose that $j \geq 3$, $a_j = 0$. We describe the case $j = 3$ and $n = 3$, the general case is done inductively.

$$\gamma = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Then

$$\gamma M(\chi) = \begin{bmatrix} a_1 & 0 \\ 0 & 1 \\ a_2 & 0 \\ 0 & 0 \\ a_2 & 0 \\ 0 & 0 \end{bmatrix}. \quad \square$$

5. The non-semisimple case

In this section we analyze lattices in those non-semisimple Lie subgroups F of $Sp(2n)$ that contain $Sp(2n-2)$. Recall that each maximal non-semisimple subgroup B of $Sp(2n)$ containing $Sp(2n-2)$, preserves a line $L \subset V^{\perp}$, where $V = \mathbb{R}^{2n-2}$ is the symplectic subspace invariant under $Sp(2n-2)$. The group B splits as semi-direct product $CH_{2n} \rtimes Sp(2n-2)$, where $CH_{2n} = H_{2n-1} \rtimes \mathbb{R}_+$ and H_{2n-1} is the $2n-1$-dimensional Heisenberg group, see §2.

Now suppose that $F = F_{\chi} \subset B$ is a Lie subgroup containing $Sp(2n-2)$. Since $Sp(2n-2)$ acts transitively on $V - 0$, the subgroup F has to be one of the following:
(a) $F = B$.
(b) $F = H_{2n-1} \rtimes Sp(2n-2)$.
(c) $F = A \times Sp(2n-2)$ where $A \subset B_0 = \mathbb{R} \rtimes \mathbb{R}_+$.

If $\Delta = F \cap Sp(2n, \mathbb{Z}) \subset F$ is a lattice then its intersection with the subgroup CH_{2n} (case (a)), H_{2n-1} (case (b)) and A (case (c)) is again a lattice (see Theorem 4.2). The first case is impossible by Lemma 2.8. In the third case the intersection $\Delta \cap Sp(2n-2)$ is a lattice as well and we are therefore reduced to the discussion in §5. This leaves us with the case (b), when $Sp(2n, \mathbb{Z}) \cap H_{2n-1}$ is a lattice. Note that there are lattices $\Delta \subset H_{2n-1} \rtimes Sp(2n-2)$ whose intersection with any conjugate of $Sp(2n-2)$ is not a lattice, we leave it to the reader to construct such examples.

Suppose now that $\chi \in X$ is a character (with the real part u and the imaginary part v) such that the closure of the orbit $\Gamma\chi$ contains the orbit $F_\chi \chi$ where $F_\chi \cong H_{2n-1} \rtimes Sp(2n-2)$ fixes a line L in $\mathrm{Span}(u, v)$. According to Remark 2.1 it suffices to consider the case $L = \mathrm{Span}(u)$. Applying an element $\gamma \in \Gamma$ we can adjust the pair (u, v) such that the vector $u = (a_1, b_1, ..., a_n, b_n)$ has no zero coordinates (see Lemma 2.6). The group H_{2n-1} acts transitively on the set of vectors $v \in \mathbb{R}^{2n}$ satisfying $\omega(u, v) = 1$. Hence we can find $h \in H_{2n-1}$ such that

$$h(v) = \frac{1}{\omega(u, v)}(...., -b_j, a_j, ...).$$

Hence $\omega_j(u, h(v)) = \omega_j(h(u), h(v)) > 0$ for each $j = 1, ..., n$. Since $\overline{\Gamma\chi}$ contains the orbit $F_\chi\chi$, there exists an element $\gamma \in \Gamma$ such that $\omega_j(\gamma(u), \gamma(v)) > 0$ for each $j = 1, ..., n$. According to Theorem 2.3 the character χ belongs to the subset $\Sigma \subset X$ of characters of abelian differentials.

6. Meromorphic differentials

THEOREM 6.1. *Suppose that $n \geq 3$ and χ is a nonzero character in $H^1(S, \mathbb{C})$ which does not satisfy either Obstruction 1 or Obstruction 2. Then there is a complex structure τ on S and a meromorphic differential α with a single simple pole on S_τ such that χ is the character of α.*

PROOF. Case A. The vectors u and v are linearly independent. The group $Sp(2n, \mathbb{R})$ acts transitively on the collection Y of pairs of vectors $u, v \in \mathbb{R}^{2n}$ such that $\omega(u, v) = 0$ and $u \wedge v \neq 0$. Thus (since $\Gamma = Sp(2n, \mathbb{Z})$ is Zariski dense in $Sp(2n, \mathbb{R})$) there exists $\gamma \in \Gamma$ such that $\chi' = \gamma\chi$ satisfies: $\omega_j(\chi') \neq 0$ for each $j = 1, ..., n$. If each $\omega_j(\chi') > 0$ then χ is the character of an abelian differential and there is nothing to prove. Hence (after relabelling j's) we get: $\omega_1(\chi') < 0$ and $\omega_j(\chi') \neq 0$, $j = 2, ..., n$. Set $\chi := \chi'$.

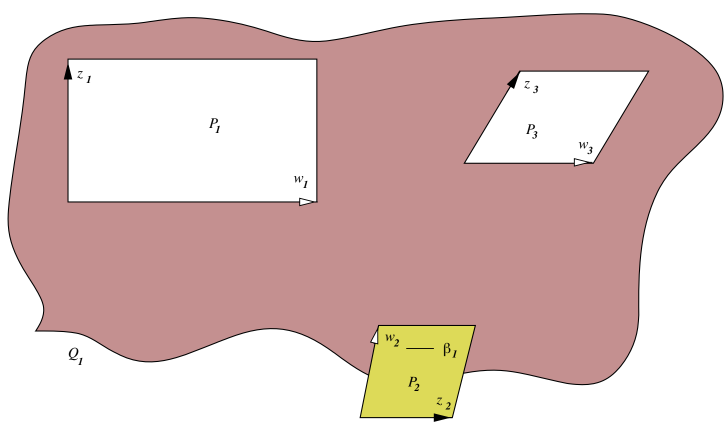


FIGURE 3

We argue similarly to the proof of Theorem 2.3. Consider the fundamental parallelogram $P_1 \subset \mathbb{C}$ for the discrete group generated by the columns z_1, w_1 of the matrix $M_1(\chi')$. Let Q_1 denote the closure of the *exterior* of P_1 in $\mathbb{S}^2$. Note that topologically Q_1 is still a parallelogram: its edges are the edges of P_1. Identifying the opposite sides of Q_1 by z_1, w_1 we get a marked torus T_1 with a standard (symplectic) system of generators x_1, y_1, branched projective structure and an orientation-preserving developing mapping to $\mathbb{S}^2$ whose holonomy is the homomorphism χ_1 which sends $x_1 \to z_1, y_1 \to w_1$. (Here we identify a vector in $\mathbb{C}$ with the corresponding translation.) Taking pull-back of the form dz on $\mathbb{C}$ we get a meromorphic differential on T_1 with the single simple pole (corresponding to the point $\infty \in Q_1$) and the period character χ_1. We now extend this to the rest of the surface S. If $j \geq 2$ is such that $\omega_j(\chi) > 0$ then similarly to the proof of Theorem 2.3 we add to T_1 the flat torus T_j obtained by identifying the sides of a fundamental parallelogram for the translation group generated by the columns of $M_j(\chi)$. If $\omega_j(\chi) < 0$ we pick a fundamental parallelogram P_j so that it is disjoint from the P_i's ($1 \leq i \leq n$, $i \neq j$). Remove the interior of P_j from Q_1 and identify the opposite sides of P_j via translations. See Figure 3.

As the outcome we get an oriented surface S and a χ-equivariant developing map to $\mathbb{S}^2$. The meromorphic differential on S is obtained via pull-back of dz from $\mathbb{C}$. Its only pole corresponds to the point on the torus T_1 which maps to ∞ under the developing map.

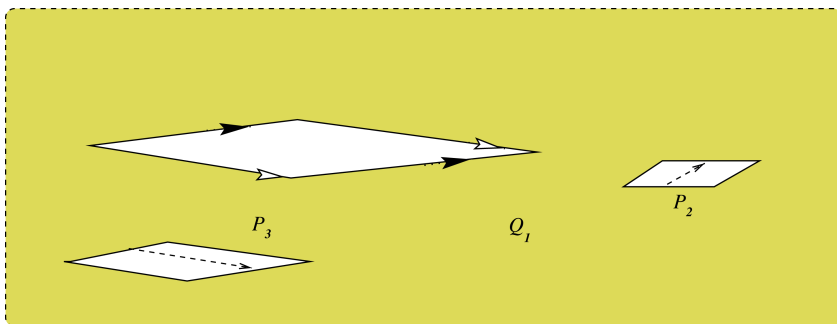


FIGURE 4

Case B. Let u and v be linearly dependent. It suffices to consider the case $u \neq 0$ (otherwise replace χ by $\sqrt{-1}\chi$). Using Zariski density of Γ in $Sp(2n, \mathbb{R})$ (the latter acts transitively on $\mathbb{R}^{2n} - 0$) choose $\gamma \in \Gamma$ such that no coordinate of $\gamma(u)$ is zero and let $\chi := \gamma\chi$. We now argue analogously to the Case A. Let z_1, w_2 denote the columns of the matrix $M_1(\chi)$. Let P_1 denote the convex hull of the set $0, z_1, w_1, z_1 + w_1$. We will think of P_1 as a degenerate parallelogram with the edges $[0, z_1], [0, w_1], [z_1, z_1 + w_1], [w_1, z_1 + w_1]$. Now cut $\mathbb{S}^2$ open along P_1 and denote the result Q_1, it is homeomorphic to a parallelogram, identification of the opposite edges via translations by z_1, w_1 yields the torus T_1. To reconstruct the rest of the surface S we choose disjoint degenerate "fundamental parallelograms" P_j for the groups generated by the translations z_j, w_j, cut Q_1 open along the P_j's ($j \geq 2$) and get S by identifying the opposite edges on each cut. See Figure 4. □

REMARK 6.2. We note that the branched projective structures σ associated to the meromorphic differentials constructed in the above theorem have the branching degree $\deg(\sigma) = 2n$.

We will next prove a lower bound on the degree of branching of the projective structures with the holonomy in the translation subgroup $\mathbb{C}$ of $PSL(2,\mathbb{C})$. This lower bound holds for all genera $n \geq 2$.

Suppose that σ is a branched projective structure with the holonomy $\rho : \pi_1(S) \to \mathbb{C} \subset PSL(2,\mathbb{C})$. We will assume that ρ is nontrivial, otherwise clearly $\deg(\sigma) \geq 2n+2$ by the Riemann–Hurwitz formula. The representation ρ lifts to a representation $\theta : \pi_1(S) \to SL(2,\mathbb{C})$ (with the image in the group of unipotent upper triangular matrices U). Let V denote the holomorphic $\mathbb{C}^2$–bundle over S associated with the representation θ. The structure σ gives rise to a holomorphic line subbundle $L \subset V$ such that

$$\deg(L) = n - 1 - \frac{\deg(\sigma)}{2}, \tag{6.1}$$

where $\deg(\sigma)$ is the degree of branching of σ (see [**GKM00**, Chapter C]). The bundle V fits into short exact sequence

$$0 \to \Lambda \to V \xrightarrow{p} \Lambda \to 0,$$

where Λ is the trivial bundle; the fibers of $\Lambda = \ker(p)$ correspond to the line in $\mathbb{C}$ fixed by the group U. Under the projectivization $\mathbb{C}^2 \to \mathbb{CP}^1$ this line projects to the point $\infty \in \mathbb{CP}^1$. Hence the developing mapping of σ does not cover ∞ iff $L \cap \ker(p) = 0$. It also follows that $L \neq \ker(p)$ (otherwise the developing mapping of σ would be constant). Therefore we get a nonzero map $p : L \to \Lambda$ by restricting the projection $p : V \to \Lambda$ to L. By the Riemann–Roch theorem, $\deg(L) \leq 0$ with the equality iff $p : L \to \Lambda$ is injective; (6.1) then implies that $\deg(\sigma) \geq 2n-2$. The equality here is attained only if the developing map of σ takes values in $\mathbb{C}$, i.e. σ is a singular Euclidean structure. In other words, if $\deg(\sigma) = 2n-2$ then the developing mapping of σ is obtained by integrating an abelian differential on S. If ρ is not the holonomy of any singular Euclidean structure then $\deg(\sigma) \geq 2n+1$. However, since ρ lifts to $SL(2,\mathbb{C})$, $\deg(\sigma)$ has to be even (see [**GKM00**, Chapter C]). We conclude that in this case $\deg(\sigma) \geq 2n$. Recall that for a representation $\rho : \pi_1(S) \to PSL(2,\mathbb{C})$, $d(\rho)$ is the least degree of branching of all projective structures on S (consistent with the orientation) with the holonomy ρ. We thus proved:

PROPOSITION 6.3. *Suppose that ρ is a representation $\rho : \pi_1(S) \to PSL(2,\mathbb{C})$ whose image is contained the translation subgroup $\mathbb{C}$ of $PSL(2,\mathbb{C})$. Then $d(\rho) \geq 2n-2$ and $d(\rho) \geq 2n$ provided that the corresponding character $\chi \in H^1(S,\mathbb{C})$ is not the period character of any abelian differential.*

Combining this proposition with Theorems 1.2 and 6.1 we obtain Corollary 1.3.

References

[BSY09] F. Bogomolov, F. Soloviev, and M. Yotov, *Realization of holomorphic one-forms and hyperbolic curves in parallelizable manifolds*, Preprint, 2009.

[CDF15] G. Calsamiglia, B. Deroin, and S. Francaviglia, *A transfer principle: from periods to isoperiodic foliations*, arXiv:1511.07635, 2015.

[Dan86] S. G. Dani, *Orbits of horospherical flows*, Duke Math. J. **53** (1986), no. 1, 177–188, DOI 10.1215/S0012-7094-86-05312-3. MR835804

[Dyn52] E. B. Dynkin, *Maximal subgroups of the classical groups* (Russian), Trudy Moskov. Mat. Obšč. **1** (1952), 39–166. MR0049903

[ECG87] R. D. Canary, D. B. A. Epstein, and P. Green, *Notes on notes of Thurston*, Analytical and geometric aspects of hyperbolic space (Coventry/Durham, 1984), London Math. Soc. Lecture Note Ser., vol. 111, Cambridge Univ. Press, Cambridge, 1987, pp. 3–92. MR903850

[GKM00] Daniel Gallo, Michael Kapovich, and Albert Marden, *The monodromy groups of Schwarzian equations on closed Riemann surfaces*, Ann. of Math. (2) **151** (2000), no. 2, 625–704, DOI 10.2307/121044. MR1765706

[Gol87] William M. Goldman, *Geometric structures on manifolds and varieties of representations*, Geometry of group representations (Boulder, CO, 1987), Contemp. Math., vol. 74, Amer. Math. Soc., Providence, RI, 1988, pp. 169–198, DOI 10.1090/conm/074/957518. MR957518

[GOV94] V. Gorbatsevich, A. Onishchik, and È. Vinberg, *Lie groups and Lie algebras, III*, Springer-Verlag, Berlin, 1994, Structure of Lie groups and Lie algebras.

[Hau20] Otto Haupt, *Ein Satz über die Abelschen Integrale 1. Gattung* (German), Math. Z. **6** (1920), no. 3-4, 219–237, DOI 10.1007/BF01199957. MR1544406

[Hej75] Dennis A. Hejhal, *Monodromy groups and linearly polymorphic functions*, Acta Math. **135** (1975), no. 1, 1–55, DOI 10.1007/BF02392015. MR0463429

[Kap95] Michael Kapovich, *On monodromy of complex projective structures*, Invent. Math. **119** (1995), no. 2, 243–265, DOI 10.1007/BF01245182. MR1312500

[Kar55] F. I. Karpelevič, *The simple subalgebras of the real Lie algebras* (Russian), Trudy Moskov. Mat. Obšč. **4** (1955), 3–112. MR0087647

[Mar89] G. A. Margulis, *Indefinite quadratic forms and unipotent flows on homogeneous spaces*, Dynamical systems and ergodic theory (Warsaw, 1986), Banach Center Publ., vol. 23, PWN, Warsaw, 1989, pp. 399–409. MR1102736

[Mar91] G. A. Margulis, *Discrete subgroups of semisimple Lie groups*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], vol. 17, Springer-Verlag, Berlin, 1991. MR1090825

[Nar92] Raghavan Narasimhan, *Compact Riemann surfaces*, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 1992. MR1172116

[Rag72] M. S. Raghunathan, *Discrete subgroups of Lie groups*, Springer-Verlag, New York-Heidelberg, 1972. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 68. MR0507234

[Rat91] Marina Ratner, *On Raghunathan's measure conjecture*, Ann. of Math. (2) **134** (1991), no. 3, 545–607, DOI 10.2307/2944357. MR1135878
1, 2 (Zürich, 1994), Birkhäuser, 1995, pp. 157–182.

[Rat95] Marina Ratner, *Interactions between ergodic theory, Lie groups, and number theory*, Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Zürich, 1994), Birkhäuser, Basel, 1995, pp. 157–182. MR1403920

[Str84] Kurt Strebel, *Quadratic differentials*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], vol. 5, Springer-Verlag, Berlin, 1984. MR743423

[Thu81] W. Thurston, *Geometry and topology of 3-manifolds*, Princeton University Lecture Notes, 1978–1981.

[Zim84] Robert J. Zimmer, *Ergodic theory and semisimple groups*, Monographs in Mathematics, vol. 81, Birkhäuser Verlag, Basel, 1984. MR776417

Department of Mathematics, University of California, 1 Shields Ave., Davis, CA 95616

Email address: kapovich@math.ucdavis.edu

Contemporary Mathematics
Volume **744**, 2020
https://doi.org/10.1090/conm/744/14991

Crossed renormalization of quadratic polynomials

Johannes Riedl and Dierk Schleicher

Abstract. Renormalization is a fundamental concept in many areas of mathematics and physics. Douady and Hubbard introduced simple renormalization in complex dynamics as a conceptual explanation why the Mandelbrot set (and many other bifurcation loci) contain countably many embedded little Mandelbrot sets. Crossed renormalization is another version of renormalization that was introduced by Curt McMullen. We study quadratic polynomials that are crossed renormalizable: we show that the locus of crossed renormalizable polynomials within the Mandelbrot set consists of countably many components, each of which is canonically homeomorphic to a (sub)limb of the Mandelbrot set; we also give a complete combinatorial description of crossed renormalizable parameters in terms of internal addresses.

We discuss similarities and differences to the theory of simple renormalization, answering a question of McMullen.

Contents

2010 *Mathematics Subject Classification.* Primary: 37F20, 37F25, 37F45.

Key words and phrases. Crossed renormalization, Mandelbrot set.

We gratefully acknowledge partial support by the ERC Advanced Grant HOLOGRAM during the preparation of this revised version of the paper.

1. Introduction

In many branches of mathematics and mathematical physics, renormalization theory plays an important role. This is especially true for complex dynamics, where powerful methods from complex analysis made a particularly deep study of renormalization possible. In the mid-1980's, Douady and Hubbard pioneered renormalization theory within complex dynamics by their theory of polynomial-like maps [**DH2**]. They introduced what is now known as "simple renormalization", and they showed that the subset in the Mandelbrot set $\mathcal{M}$ consisting of simple renormalizable parameters consists of countably many homeomorphic copies of $\mathcal{M}$ embedded within $\mathcal{M}$ (finitely many copies for each renormalization period); in fact, they discovered that renormalization is the reason for the existence of all the embedded copies of $\mathcal{M}$ within itself (except that, strictly speaking, the image of the root of $\mathcal{M}$ may fail to be renormalizable). Later, McMullen [**McM**, Chapter 7], in his study of renormalization theory within complex dynamics, discovered a different kind of renormalization called "crossed renormalization", and asked for a description of crossed renormalization similar to the known description of simple renormalization. This is the purpose of the present paper. Our main result is the following (terminology and background will be explained in Section 2).

Theorem 1.1 (The locus of crossed renormalizations).
For every period $n \geq 2$, the locus of crossed n-renormalization within the Mandelbrot set consists of countably infinitely many connected subsets of $\mathcal{M}$, each of which is canonically homeomorphic to a limb of the Mandelbrot set. More precisely, all parameters that are crossed n-renormalizable around fixed points are contained in p/qn-limbs of $\mathcal{M}$, for integers $q > p > 0$ so that p and qn are coprime, and so that the corresponding subset within the p/qn-limb is canonically homeomorphic to the p/q-limb of $\mathcal{M}$. All parameters that are crossed renormalizable around periodic points are simple renormalizable and are contained in the corresponding sublimbs of the "little Mandelbrot sets" coming from simple renormalization.

We will make more precise statements in Sections 3 and 4 where we describe how to find these embedded limbs: in Section 3, we will describe in detail the special case that the crossed renormalization is around a fixed point, and we will show in Section 4 that the discussion of the general case, that the renormalization is around a periodic point, can readily be reduced to the fixed point case. A first description of crossed renormalization will be given along with the fundamental dynamical construction in Section 3.1. In Section 3.2, we will describe the crossed renormalization locus by chopping off subsets of the Mandelbrot set until a component of the renormalization locus remains. In Section 3.3, we show that connected components of the loci of crossed renormalization are homeomorphic to limbs of the Mandelbrot set. The proof that our construction captures every case of crossed renormalization will be given in Section 3.4. An informal model for the dynamics of crossed renormalizable polynomials can be found in Section 3.5. Finally, we show how to tell which quadratic polynomials are crossed renormalizable in terms of internal addresses (Section 3.6) and in terms of tableaux of puzzles (Section 3.7).

We describe crossed renormalization for quadratic polynomials, normalized as $z \mapsto z^2 + c$. The results on the structure of renormalization, simple and crossed, are quite analogous for unicritical polynomials $z \mapsto z^d + c$ for degrees $d \geq 2$.

A preliminary version of this paper appeared informally in the Proceedings of the Kyoto workshop in holomorphic dynamics in October 1997 [**RS**], based on the first author's Master Thesis. This paper, which constitutes the formal publication of that preprint, has been significantly revised and updated.

Acknowledgement. We would like to thank Shunsuke Morosawa for having organized the aforementioned workshop, and Shizuo Nakane and Mitsuhiro Shishikura for having made the participation possible. Far more recently, we would like to thank the anonymous referee sincerely for their extremely valuable and constructive comments and suggestions.

2. Background on quadratic polynomials

In this section we describe necessary background from complex dynamics. The proofs for the statements can be found in [**DH1**, **M2**] or [**S5**].

2.1. The Mandelbrot set and Julia sets. In this paper, we will discuss exclusively quadratic polynomials; they can all be affinely conjugated to the form $P_c\colon z \mapsto z^2 + c$ for a unique complex parameter c. The *filled-in Julia set* K_c for the polynomial P_c is the set of points z in the dynamic plane such that the iterates of z do not escape to infinity, i.e. the sequence $z, P_c(z), P_c^{\circ 2}(z), \dots$ is bounded. The *Julia set* J_c is the boundary of the filled-in Julia set. Both sets are compact and K_c is full (which means that the complement is connected). The dynamics of a rational map is determined to a large extent by the critical points of the map (those points where the dynamics fails to be locally injective) and their forward orbits. One example of this observation is that the filled-in Julia set of a polynomial is connected if and only if it contains all the critical points in $\mathbb{C}$ of the polynomial, i.e. if all the critical orbits in $\mathbb{C}$ are bounded.

For our polynomials $P_c(z) = z^2 + c$, the only critical point in $\mathbb{C}$ is 0. Therefore, a filled-in Julia set K_c is connected iff 0 does not escape to ∞ under iteration.

The *Mandelbrot set* $\mathcal{M}$ is the set of all parameters c such that the filled-in Julia set K_c (or equivalently the Julia set J_c) is connected, so it is often referred to as the *quadratic connectedness locus*. By fundamental work of Douady and Hubbard [**DH0**, **DH1**], it is known to be compact, connected and full.

For every period $n \geq 1$, there are open sets in parameter space for which the polynomial P_c has an attracting periodic orbit of period n. These sets are contained in the Mandelbrot set, and their connected components are called *hyperbolic components* of the Mandelbrot set [**DH1**, **M3**, **S2**]. For every fixed period, their number is finite. Every hyperbolic component is conformally parametrized by the multiplier of the attracting orbit: the multiplier map supplies a natural biholomorphic map between the component and the open unit disk, and it extends as a homeomorphism to the closures. Every hyperbolic component has a unique *center* and a unique *root*: these are the points where the (extended) multiplier map takes values 0 and +1, respectively.

2.2. Dynamic rays, parameter rays, and equipotentials. Let us consider a polynomial P_c for some parameter $c \in \mathcal{M}$. The dynamics outside the filled-in Julia set can conveniently be described by *dynamic rays* (also known as *external rays*) and *equipotentials*, which are a dynamic variant of polar coordinates. Since K_c is full, there is a conformal isomorphism $\varphi_c\colon \overline{\mathbb{C}} \setminus K_c \to \overline{\mathbb{C}} \setminus \overline{\mathbb{D}}$ fixing ∞; it is unique up to rotation and can be fixed so that $\lim_{z\to\infty} \varphi_c(z)/z$ is real positive. In fact, since our polynomials are normalized so that their leading coefficient is 1, the limit will be equal to 1. The map φ_c conjugates the dynamics in $\overline{\mathbb{C}} \setminus K_c$ to the dynamics of $z \mapsto z^2$ in $\overline{\mathbb{C}} \setminus \overline{\mathbb{D}}$: we have $(\varphi_c(z))^2 = \varphi(z^2 + c)$.

For every $\vartheta \in \mathbb{S}^1 = \mathbb{R}/\mathbb{Z}$, the set $R_c(\vartheta) := \varphi_c^{-1}(\{e^r \cdot e^{2\pi i\vartheta} \colon 0 < r < \infty\})$ is called the *dynamic ray of P_c at angle ϑ*. Note that external angles are counted with respect to full turns (for instance, angle 1/4 denotes a quarter of a turn, i.e. a right angle). By the conjugation property we have

$$P_c(R_c(\vartheta)) = R_c(2\vartheta) \ .$$

In our parametrization, a dynamic ray is periodic if its angle is rational with odd denominator when written in lowest terms; it is strictly preperiodic if the angle is rational with even denominator; and it has an infinite forward orbit if the angle is irrational. (These statements concern the rays as sets; of course each point on a ray escapes.)

For any $\sigma \in (0, \infty)$, the set $E_\sigma := \varphi_c^{-1}(\{e^\sigma \cdot e^{2\pi i\vartheta} \colon \vartheta \in \mathbb{S}^1\})$ is called the *equipotential of P_c at potential σ*. The conjugation property yields

$$P_c(E_\sigma) = E_{2\sigma} \ .$$

The dynamic rays together with the equipotentials form a coordinate system in $\mathbb{C} \setminus K_c$ in which the dynamics is simply doubling of external angles and potentials. It is often useful, but not always possible, to extend this coordinate system to the Julia set. A dynamic ray at angle ϑ is said to *land* at a point z of the Julia set if

$$\lim_{\sigma \to 0} \varphi_c^{-1}(e^\sigma \cdot e^{2\pi i\vartheta}) = z.$$

In general, not every dynamic ray needs to land; its limit set is always a connected subset of the Julia set. However every dynamic ray at a rational angle lands at a periodic or preperiodic point of the Julia set; conversely, every repelling periodic or preperiodic point is the landing point of some dynamic rays with rational angles, and all the rays landing at the same point have the same periods and preperiods [**M1**, §18]. A dynamic *ray pair* is the union of two dynamic rays that land at the same point, together with their landing point; if it consists of the two rays at angles ϑ and ϑ', we denote this ray pair simply by $\langle\vartheta, \vartheta'\rangle$. A ray pair is *characteristic* if it separates the critical value from the critical point, and if moreover this ray pair has no ray pair on its forward orbit that separates it from the critical value.

It will be necessary later to use dynamic rays not only for parameters within the Mandelbrot set but also for $c \notin \mathcal{M}$: for every parameter c there exists a holomorphic injective mapping φ_c defined in a neighborhood of infinity that conjugates the dynamics of P_c to $z \mapsto z^2$ near infinity. This defines dynamic rays above a certain potential σ. If a dynamic ray at angle ϑ is defined above potential σ and does not contain the critical value, then the preimage under P_c of this ray defines the two dynamic rays at angles $\vartheta/2$ and $(\vartheta+1)/2$ above potential $\sigma/2$. Repeating this procedure countably often, dynamic rays are defined for all positive potentials, for all but possibly countably many angles. Observe that the definition of potentials extends naturally to all of $\mathbb{C}$: sufficiently close to ∞ the potential of z is defined as $\log|\varphi_c(z)|$, and we can recursively define the potential of z as 1/2 times the potential of $P_c(z)$ for all points in $\mathbb{C} \setminus K_c$. Finally, points in K_c have potential 0.

Since the Mandelbrot set is compact, connected and full, there is a Riemann map $\Phi\colon \overline{\mathbb{C}}\setminus\mathcal{M} \to \overline{\mathbb{C}}\setminus\overline{\mathbb{D}}$, and it can be normalized uniquely so that $\Phi(c)/c \to 1$. Using this Riemann map, can define external rays and equipotentials of the Mandelbrot set as well. In order to distinguish these rays from the rays in dynamical planes, we call them *parameter rays*. Again, all parameter rays at rational angles land [**DH1**, **M3**, **S2**]. More precisely, for every period $n \geq 1$, exactly two parameter rays at angles of period n land together, and the landing point is the root of a hyperbolic component of period n; conversely, the root of every hyperbolic component of period n is exactly the landing point of two parameter rays of period n (for period $n = 1$, we count the two parameter rays at angles $\vartheta = 0$ and $\vartheta = 1$ separately). Parameter rays at preperiodic angles land at parameters for which the critical orbit is strictly preperiodic; such parameters are known as Misiurewicz–Thurston points [**DH1**, **S2**]. The number of parameter rays landing at any given Misiurewicz–Thurston point is positive and finite. We define *parameter ray pairs* in analogy to dynamic ray pairs and denote the parameter ray at angle ϑ by $R_{\mathcal{M}}(\vartheta)$.

We mentioned earlier that every hyperbolic component W of period n has a natural conformal isomorphism $\mu\colon W \to \mathbb{D}$ that extends as a homeomorphism $\mu\colon \overline{W} \to \overline{\mathbb{D}}$. This yields a natural parametrization of ∂W by $\mathbb{S}^1$, setting $c = \mu^{-1}(e^{2\pi i\vartheta})$ for $\vartheta \in \mathbb{S}^1 = \mathbb{R}/\mathbb{Z}$. Every boundary point at a rational internal angle $\vartheta = p/q \neq 0$ is a *bifurcation point*: at this point, a hyperbolic component of period qn is attached. The bifurcation point is the root of the qn-periodic component, and the pair of periodic parameter rays landing at this root (with angles of period qn) separates the component of period qn from the component of period n and the origin. The open region that is separated from the origin by this parameter ray pair is called the *p/q-subwake of W*. The intersection of this wake with $\mathcal{M}$ is the *p/q-sublimb of W*. The landing point of the two bounding parameter rays is the *root* of sublimb and subwake. (By our convention, the subwake is open in $\mathbb{C}$ and the sublimb is relatively open in $\mathcal{M}$, so both do not contain the root point). Subwake and sublimb at internal angle p/q of the unique hyperbolic component of period 1 are called the *p/q-wake* and *p/q-limb* of the Mandelbrot set; the p/q-limb will be called $\mathcal{M}_{p/q}$. The *root* of the limb or wake is the common boundary point of limb or wake with the period 1 hyperbolic component of $\mathcal{M}$; we will denote it $c_{p/q}$.

Now let $c \in \mathbb{C} \setminus [1/4, \infty)$ be a parameter (not necessarily in $\mathcal{M}$). Then P_c has exactly two fixed points in $\mathbb{C}$. One of them is the landing point of the dynamic ray at angle $\vartheta = 0$; this fixed point is called the β-fixed point of P_c. The other fixed point is called the α-fixed point; it may be attracting, indifferent, or repelling, and it may or may not be the landing point of periodic dynamic rays. If it is, the period of these dynamic rays must be some finite number $q \geq 2$ (because the only ray of period 1 is already taken). The combinatorial rotation number of the dynamics of these q rays is then p/q for some integer p coprime to q. It turns out that the subset of $\mathbb{C}$ for which the α-fixed point is repelling and the landing point of q rays with combinatorial rotation number p/q is exactly the p/q-wake of $\mathcal{M}$ as defined above [**M3**]. For every parameter $c \in \mathbb{C}$, exactly one of the following holds:

- c is in the closure of the hyperbolic component of period 1, and the α-fixed point is attracting or indifferent;
- c is in the p/q-wake of $\mathcal{M}$ for a unique $p/q \in (0, 1)$ and the α-fixed point is repelling and the landing point of q dynamic rays with combinatorial rotation number p/q;
- $c \notin \mathcal{M}$ is on the boundary of the p/q-wake for a unique $p/q \in (0, 1)$ and the α-fixed point is repelling but not the landing point of any ray;
- $c \in (1/4, \infty)$;

- $c \notin \mathcal{M}$ is on one of uncountably many parameter rays outside of the is closures of all wakes and α is the landing point of infinitely many dynamic rays, all of them non-periodic.

The last case is the Cantor set analogue to Siegel disks or Cremer points of period 1, and all the rays that, for Siegel disks, "want to" land on the boundary at the Siegel disk all land at the α fixed point in this case. Since this case might be lesser known, we outline some of the key arguments. The first observation is that in the disconnected, hence Cantor set case, every ray must land (unless it is one of the countably many that contain a point on the backwards orbit of the critical value). The next observation is that the parameter rays $R_{\mathcal{M}}(\vartheta)$ outside of closures of all wakes occur exactly for those angles ϑ for which $2^k\vartheta \in (\vartheta/2, (\vartheta+1)/2)$ for all $k \geq 0$: in the dynamics of such a parameter c, the critical value c is on the dynamic ray $R_c(\vartheta)$, and the two preimage rays $R_c(\vartheta/2)$ and $R_c((\vartheta+1)/2)$ both crash into the critical point and divide $\mathbb{C}$ into two components. The condition on $2^k\vartheta$ expresses the fact that the entire critical orbit is contained in the same component (it can also be re-interpreted as saying that the kneading sequence associated to ϑ is a constant sequence [**S4**]). Now consider the set of angles φ so that $2^k\varphi \in [\vartheta/2, (\vartheta+1)/2]$ for all $k \geq 0$. This set is a Cantor set (this follows easily from the condition on the orbit $2^k\vartheta$), and every such dynamic ray $R_c(\varphi)$ must then land at the α fixed point, except the countably many rays that do not land at all.

In all cases when the α-fixed point is the landing point of at least two rays, it must disconnect the filled-in Julia set K_c (provided it is connected in the first place). The β-fixed point never disconnects K_c. We say that a Misiurewicz–Thurston point is of *α-type* or of *β-type* if the critical orbit eventually lands at the α-fixed point, respectively at the β-fixed point.

2.3. Polynomial-like maps. A *polynomial-like map* $f\colon U \to V$ is a proper holomorphic map f between two bounded, open, connected and simply connected domains $U, V \subset \mathbb{C}$ such that $\overline{U} \subset V$; these maps were introduced by Douady and Hubbard in [**DH2**]. Such a map has a mapping degree $d \geq 1$. If this degree is 2, we call it a *quadratic-like map.* Every polynomial p of degree $d \geq 2$ becomes a polynomial-like map of degree d when V is a sufficiently large disk centered at 0 and $U = p^{-1}(V)$. But often the dynamics of a high iterate of a polynomial, which itself is a polynomial of large degree, can be understood more easily by restricting it to an appropriate subset on which the dynamics is polynomial-like of much smaller degree.

The filled-in Julia set $K(f)$ of a polynomial-like mapping $f\colon U \to V$ is the set of all points $z \in U$ that never leave U under iteration of f. The Julia set $J(f)$ of f is the boundary of $K(f)$. As for actual polynomials, these sets are connected iff all the critical points of f are contained in $K(f)$.

An important statement is the *Straightening Theorem* of Douady and Hubbard [**DH2**, Theorem 1]. The definition of quasiconformal mappings and further details can also be found there; a more recent textbook reference is [**BF**].

Theorem 2.1 (The Straightening Theorem).
Let $f\colon U \to V$ be a polynomial-like map of degree $d \geq 2$. Then there exists a polynomial P of degree d such that f and P are hybrid equivalent, i.e., possibly after shrinking V to an appropriate neighborhood of $K(f)$, and U so that the property of a polynomial-like map is maintained, there is a neighborhood V_P of $K(P)$ and a

quasiconformal homeomorphism $\varphi\colon V \to V_P$ *such that*

$$\varphi \circ f \circ \varphi^{-1} = P \ \text{ on } P^{-1}(V_P) \quad \text{ and } \quad \frac{\partial \varphi}{\partial \overline{z}} = 0 \text{ almost everywhere on } K(f) \ .$$

If $K(f)$ *is connected, the polynomial* P *is unique up to affine conjugation.* □

For quadratic polynomials in the normalization $z^2 + c$, every parameter c represents its own affine conjugation class, so there exists a straightening map that sends any quadratic-like map f with connected Julia set to a well defined parameter $\chi(f) \in \mathcal{M}$. If $K(f)$ is not connected, then there is no hope for uniqueness: all quadratic polynomials with disconnected Julia sets are hybrid equivalent. One can still define a straightening map χ in the disconnected case, but it depends on certain choices; one way to specify these choices is called a *tubing* [**DH2**, Chapter II] (we will not specify the precise definition of a tubing here). In the disconnected case χ takes values in $\mathbb{C} \setminus \mathcal{M}$. This is relevant in particular for a family of polynomial-like maps, as follows:

Definition 2.2 (Analytic family of polynomial-like mappings).
For an open set $\Lambda \subset \mathbb{C}$, consider a family $(f_\lambda\colon U_\lambda \to V_\lambda)_{\lambda\in\Lambda}$ of polynomial-like maps. Let

$$\begin{aligned} &\mathcal{U} := \{(\lambda, z) \in \Lambda \times \mathbb{C}\colon \lambda \in \Lambda, z \in U_\lambda\} \ , \\ &\mathcal{V} := \{(\lambda, z) \in \Lambda \times \mathbb{C}\colon \lambda \in \Lambda, z \in V_\lambda\} \ , \quad \text{and} \\ &F\colon \mathcal{U} \to \mathcal{V}, \ F(\lambda, z) = (\lambda, f_\lambda(z)) \ . \end{aligned}$$

Then $(f_\lambda)_{\lambda\in\Lambda}$ is called an *analytic family of polynomial-like mappings* if the following three conditions are satisfied:

(1) $\mathcal{U}$ and $\mathcal{V}$ are homeomorphic over Λ to $\Lambda \times \mathbb{D}$,
(2) the projection from the closure of $\mathcal{U}$ in $\mathcal{V}$ to Λ is proper, and
(3) the mapping $F\colon \mathcal{U} \to \mathcal{V}$ is complex-analytic and proper.

Lemma 2.3 (Constructing an analytic family).
Let $U_0 \subset\subset V_0 \subset \mathbb{C}$ *be bounded Jordan domains,* $\gamma_U : \mathbb{D} \times \partial U_0 \to \mathbb{C}$ *and* $\gamma_V : \mathbb{D} \times \partial V_0 \to \mathbb{C}$ *continuous such that* $\gamma_U(0, .) = \mathrm{id}|_{\partial U_0}$, $\gamma_V(0, .) = \mathrm{id}|_{\partial V_0}$ *and* $\gamma_U(\lambda, .)$, $\gamma_V(\lambda, .)$ *are injective for all* $\lambda \in \mathbb{D}$. *Let* U_λ *and* V_λ *be the bounded components of* $\mathbb{C} \setminus \gamma_U(\lambda, \partial U_0)$ *and* $\mathbb{C} \setminus \gamma_V(\lambda, \partial V_0)$.

Denote by $\mathcal{U}$ *and* $\mathcal{V}$ *the sets of all* $(\lambda, z) \in \mathbb{D} \times \mathbb{C}$ *such that* $z \in U_\lambda$ *and* $z \in V_\lambda$ *respectively. If*

$$F\colon \mathcal{U} \to \mathcal{V}, \ F(\lambda, z) = (\lambda, f_\lambda(z))$$

is a complex-analytic mapping with mapping degree $d > 0$, *then* $(f_\lambda)_{\lambda\in\mathbb{D}}$ *is an analytic family of polynomial-like mappings.*

Proof. The first condition in Definition 2.2 is satisfied because we can choose Riemann mappings $U_\lambda \to \mathbb{D}$ and $V_\lambda \to \mathbb{D}$ that depend continuously on λ. The second condition also follows from continuous dependence on ∂U_λ on λ. The last one is obvious. □

For quadratic-like mappings with disconnected Julia set the straightening can be done in a way such that the following theorem holds [**DH2**, Proposition 14 and Theorem 4]:

THEOREM 2.4 (Properties of straightening map).
For every analytic family $(f_\lambda)_{\lambda\in\Lambda}$ *of quadratic-like maps where* Λ *is contractible, the straightening map* $\chi\colon \Lambda \to \mathbb{C}$ *can be defined so as to be continuous everywhere, and analytic on* $\chi^{-1}(\mathcal{M}\setminus\partial\mathcal{M})$.

On any given component of $\chi^{-1}(\mathcal{M}\setminus\partial\mathcal{M})$ *on which the map* χ *is not constant, it is an open map and has a local mapping degree; this local degree is* 1 *except at a discrete set of points.*

PROOF. By [**DH2**, Proposition 8], every analytic family of polynomial-like maps with contractible Λ has a tubing (which involves some choices), and this makes χ well defined for all Λ. By [**DH2**, Theorem 2], this map χ is continuous, and even analytic over $\Lambda\setminus\chi^{-1}(\mathcal{M}\setminus\partial\mathcal{M})$. Of course, by the straightening theorem, the set $\chi^{-1}(\mathcal{M})$ and the value of χ on this set do not depend on any choices.

If Λ is connected and $\chi\colon \Lambda\to\mathbb{C}$ is not constant, then χ is *topologically holomorphic over* $\mathcal{M}$: "it has the same topological properties it would have if it were holomorphic" [**DH2**, Chapter IV]. In particular, $\chi\colon \chi^{-1}(\mathcal{M}) \to \chi(\chi^{-1}(\mathcal{M}))$ has a local mapping degree, and this degree is 1 except over a closed discrete set of points. □

REMARK 2.5. For us, the set Λ of parameters is a subset of $\mathbb{C}$, so we have a complex one-dimensional family of polynomial-like maps. Many statements carry over to the case that Λ is a complex manifold; only the second half of Theorem 2.4, concerning the local mapping degree, requires Λ to be complex one-dimensional.

2.4. Renormalization. A quadratic polynomial P_c is called *n-renormalizable* if there are neighborhoods U, V of the critical point such that the restriction $P_c^{\circ n}\colon U\to V$ is a quadratic-like map with connected filled-in Julia set K. This set K is often referred to as the *little filled-in Julia set* of the renormalization, and its boundary is the *little Julia set*. Obviously, $P_c^{\circ n}(K) = K$.

It is known [**McM**, Theorem 7.3] that K can meet any $P_c^{\circ j}(K)$ (for $1 \le j \le n-1$) at most at a single point p, and any $P_c^{\circ j}(K)$ that intersects K does so at the same point p. If K contains such a point p, it is necessarily periodic and repelling and its period strictly divides n, so that p is a fixed point of $P_c^{\circ n}$ within K. Since the little Julia set is connected, the straightening theorem turns it into the Julia set of a polynomial $\tilde P\in\mathcal{M}$, and the two fixed points of $P_c^{\circ n}$ in K correspond to the two fixed points of $\tilde P$. It makes thus sense to speak of the α- and β-fixed points of K, in such a way that the α-fixed point of K disconnects K, while the β-fixed point does not. Since K can never meet any $P_c^{\circ j}(K)$ at α and at β simultaneously, we have the following distinction [**McM**].

DEFINITION 2.6 (Types of renormalization).
We say that an n-renormalization of a quadratic polynomial P_c with little Julia set K is of

- **disjoint type** if $K\cap P_c^{\circ j}(K)=\emptyset$ for $1\le j\le n-1$;
- **β-type** if K meets $P_c^{\circ j}(K)$ only at the β-fixed point of K;
- **α-type** if K meets $P_c^{\circ j}(K)$ only at the α-fixed point of K.

The first two types are also known as *simple renormalizations*, while the last type is known as *crossed renormalization*. If the little Julia set does meet some of its forward images (in the α- and β-cases), then the renormalization is called *immediate* if the intersection point is a fixed point of P_c.

The term "crossed renormalization" indicates that the intersection point of K and $p^{\circ j}(K)$ (the α-fixed point of K) not only disconnects K as well as $p^{\circ j}(K)$, but that K and $p^{\circ j}(K)$ (and possibly $p^{\circ 2j}(K), \dots$) intersect at α in such a way that they cross each other; this follows from the fact that all the rays that land at the α-fixed point of K are on a single cycle and are permuted transitively.

Examples of simple renormalizations are shown in Figure 4 (disjoint case) and in Figure 1 (β-case). An example of a crossed renormalization is given in Figure 2. Further examples are shown and discussed in [**McM**, Chapter 7].

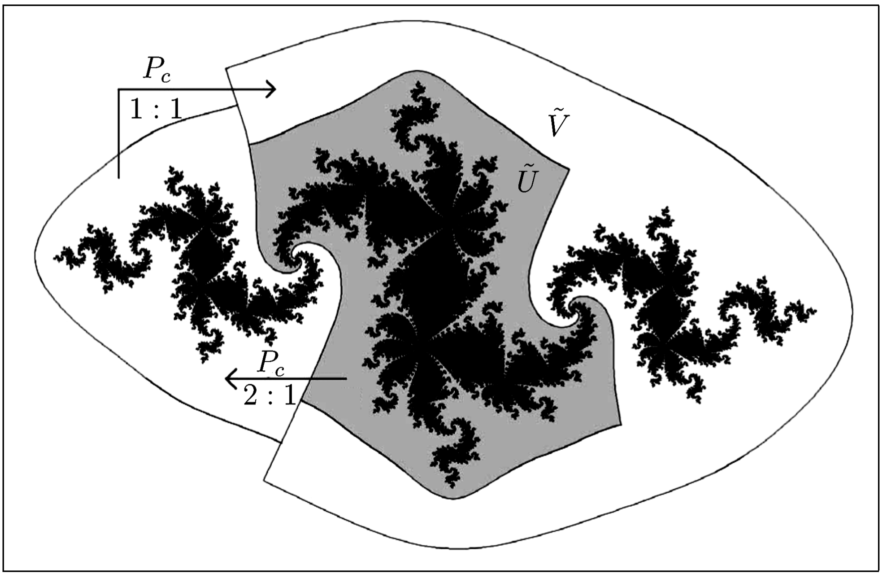


FIGURE 1. An example of a simple renormalization of β-type, with period $n=2$. The domains U and V are appropriate neighborhoods of $\tilde{U}$ and $\tilde{V}$.

A well known folklore result that goes back to the work of Douady and Hubbard [**DH2**] states that for every $n \geq 2$ the locus of disjoint-type n-renormalization consists of finitely many connected components, each of which is homeomorphic to the entire Mandelbrot set. For β-type, every component of the renormalization locus is homeomorphic to the entire Mandelbrot set without its root $c = 1/4$, and so that the homeomorphism can be extended from the closure of the renormalization locus to the entire Mandelbrot set. In both cases, a canonical homeomorphism from such a connected component to $\mathcal{M}$ is given by the straightening map χ for the quadratic-like maps arising in the renormalization process. This construction is described in Haïssinsky [**Ha**].

For every hyperbolic component W of $\mathcal{M}$ of period $n \geq 2$, there exists a unique connected component of the simple n-renormalization locus that contains W, so that χ sends the closure of the renormalization component onto $\mathcal{M}$, and in particular the component W onto the main cardioid of $\mathcal{M}$. Conversely, every connected component of the simple n-renormalization locus (in the disjoint and the β case) is based at a hyperbolic component of period n in this way. Any homeomorphism from $\mathcal{M}$ onto (the closure of) a component of the simple n-renormalization locus given as the inverse of the straightening map is called a *tuning map* of period n. This way, simple renormalization and tuning are inverse maps to each other.

Hyperbolic components of $\mathcal{M}$ come in two kinds, *primitive* or *satellite* components, depending on whether or not the root of the component is on the boundary of a hyperbolic component of lower period: if it is not, then the component has the characteristic cusp like the main component of $\mathcal{M}$, and it is called a *primitive component*; if it is, then the component has smooth boundary and is called a *satellite component*. Primitive components correspond to a renormalization component of disjoint type, while satellite components correspond to β-type. In particular, satellite components bifurcating immediately from the main cardioid of $\mathcal{M}$ correspond to immediate β-renormalization.

Further properties of simple renormalization will be mentioned throughout the paper in comparison with crossed renormalization.

Since quadratic polynomials that are crossed renormalizable around periodic points of period $m > 1$ are always simple m-renormalizable (Theorem 4.1), the focus of most research has been on simple renormalization. However, there are results on the Mandelbrot set that depend specifically on crossed renormalization: for instance, core entropy of quadratic polynomials, as introduced by William Thurston [**T**], can be defined as the limit superior of the exponential growth rate of the number of preimages of the critical point that separate the α fixed point from its negative. The natural question whether this limes superior is actually a limit depends on whether the polynomial is renormalizable; and here one has to take into account simple as well as crossed renormalization [**DS2**, Section 6].

3. Crossed renormalization: the immediate case

If a quadratic polynomial is crossed n-renormalizable, then the little Julia set meets some of its images at a periodic point which separates the little Julia sets. Recall that this renormalization is called *immediate* if this periodic point is a fixed point.

Fix an integer $n \geq 2$ and consider two positive integers p, q with $0 < p < qn$ such that p and qn are coprime. Within $\mathcal{M}_{p/nq}$ (the p/qn-limb of $\mathcal{M}$), we will be interested in the set of parameters that are immediately n-renormalizable of crossed type; we will show below that every immediately n-renormalizable parameter arises in this way.

We begin the study by constructing, for every parameter in the p/qn-wake of $\mathcal{M}$, a particular quadratic-like map (Section 3.1). The polynomial is crossed renormalizable if the Julia set of this quadratic-like map is connected. We will denote the locus of such polynomials by $C^n_{p,q}$. The straightening map defines a canonical map from $C^n_{p,q}$ to $\mathcal{M}_{p/q}$ which turns out to be a homeomorphism (Section 3.3). We show that these sets $C^n_{p,q}$ contain every immediately n-renormalizable parameter of crossed type in $\mathcal{M}$ (Section 3.4) and that $C^n_{p,q}$ can be obtained from the entire p/qn-limb by cutting off subsets bounded by certain pairs of preperiodic parameter rays landing at Misiurewicz–Thurston points (Section 3.2). We briefly sketch a description of the dynamics of crossed renormalizable polynomials in Section 3.5. Finally, we will show how to tell whether a parameter in $C^n_{p,q}$ is immediately n-renormalizable of crossed type in terms of internal addresses (Section 3.6) and in terms of puzzles and tableaux (Section 3.7). The general (non-immediate) case of crossed renormalization will be dealt with in Section 4.

3.1. The principal construction. For every parameter c in the p/qn-wake of the Mandelbrot set, we will now construct a quadratic-like map by restricting the n-th iterate of the original polynomial to an appropriate dynamically defined subset. We do not require $c \in \mathcal{M}$.

Let $\sigma_0 \geq 0$ be the potential of the critical value (with $\sigma_0 = 0$ iff $c \in \mathcal{M}$). Any equipotential $\sigma > \sigma_0$ bounds a topological disk that we will call $V(\sigma)$. It contains the filled-in Julia set $K(P_c)$ of P_c. From now on, fix a potential $\sigma > 2^n\sigma_0$.

For all parameters c in the p/qn-wake, the α-fixed point is the landing point of exactly qn dynamic rays, and these are permuted transitively by the dynamics of the polynomial [**M3**]. This permutation has combinatorial rotation number p/qn, i.e. every ray jumps over $p-1$ rays onto its image ray, counting counterclockwise. Similarly, the point $-\alpha$ is the landing point of equally many preperiodic rays in a configuration that is symmetric to the rays landing at α, and the polynomial maps $-\alpha$ with its rays onto α with its rays. None of these rays contains the critical point (or the rays would not exist), and obviously α or $-\alpha$ cannot be the critical point, either. In particular, $\alpha \neq -\alpha$.

The qn rays landing at α, together with the qn rays landing at $-\alpha$, cut $V(\sigma)$ into $2qn-1$ components in a symmetric way (see Figure 2). The component containing the critical point is itself symmetric and meets both α and $-\alpha$ on its boundary; call its closure Y_0. The closures of the remaining $qn-1$ components at α will be called sectors and labeled $Y_1, \ldots, Y_{qn-1}$ ordered by the dynamics, i.e. such that $Y_j \cap V(\sigma/2)$ is mapped onto Y_{j+1} for $j = 0, 1, \ldots, qn-2$. Finally, the closures of the remaining $qn-1$ components at $-\alpha$ will also be called sectors and labeled $Z_1, \ldots, Z_{qn-1}$ such that $Z_j = -Y_j$. Then $Z_j \cap V(\sigma/2)$ maps to Y_{j+1} for $j \leq qn-2$. Since they have the same image as the Y_j and the global degree of P_c is two, the restriction of P_c onto any Y_j or Z_j is injective for $j \neq 0$. Let $Z := Z_1 \cup Z_2 \cup \ldots \cup Z_{qn-1}$. Then the common image of Y_{qn-1} and Z_{qn-1}, restricted to $V(\sigma/2)$, is $Y_0 \cup Z$; again, the map is injective separately on Y_{qn-1} and on Z_{qn-1}. Finally, $Y_0 \cap V(\sigma/2)$ contains the critical point and maps onto Y_1 in a two-to-one fashion.

All the Y_j and Z_j together cover $V(\sigma)$, and they are disjoint except at their boundaries. The restriction of P_c to $V(\sigma/2)$ is a quadratic-like map with range $V(\sigma)$. We will now identify a smaller subset of $V(\sigma/2^n)$ such that the n-th iterate of P_c is a quadratic-like map with range contained in $V(\sigma)$.

As a first step define

$$\tilde{U} := \left(Y_0 \cup \bigcup_{j=1}^{q-1} (Y_{jn} \cup Z_{jn}) \right) \cap V(\sigma/2^n) \quad \text{and} \quad \tilde{V} := Y_0 \cup \bigcup_{j=1}^{q-1} Y_{jn} \cup \bigcup_{j=1}^{qn-1} Z_j$$

(see Figure 2). Then $P_c^{\circ n} \colon \operatorname{int}(\tilde{U}) \to \operatorname{int}(\tilde{V})$ is a proper map of degree 2. However, there are two problems: $\operatorname{int}(\tilde{U})$ and $\operatorname{int}(\tilde{V})$ are disconnected; and $\tilde{U}$ and $\tilde{V}$ have common boundary points along entire ray segments. These two problems and their solutions are quite standard (compare for example Milnor [**M2**]).

The first problem can be cured by adding a small disk around α and $-\alpha$ to $\tilde{U}$: since α is repelling, we have $|P_c'(\alpha)| > 1$ and thus there is an $\varepsilon > 0$ such that $D := \{z \colon |z - \alpha| < \varepsilon\}$ has the property that $P_c^{\circ n} \colon D \to P_c^{\circ n}(D)$ is a conformal isomorphism with $P_c^{\circ n}(D) \supset \overline{D}$. Then $\operatorname{int}(\tilde{U} \cup D \cup (-D))$ is connected and it is mapped two-to-one onto its image by $P_c^{\circ n}$ provided ε is small enough.

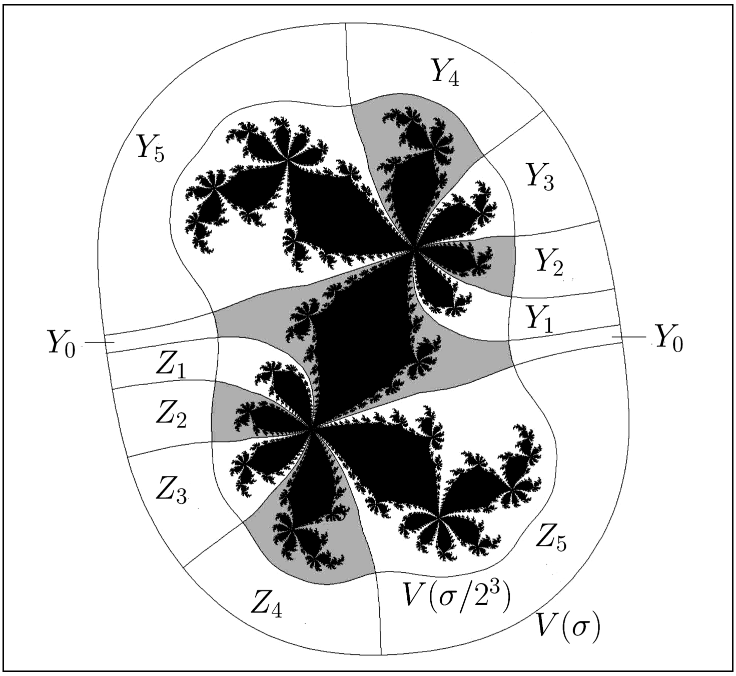


FIGURE 2. An example of crossed renormalization with $n = 2$ and $p/q = 1/3$. The little Julia set is the set of points with orbits that remain in the shaded domain (it is homeomorphic to the well known "Douady rabbit"). This figure also illustrates the construction of the quadratic-like maps: the set $\tilde{U}$ is shaded, and $\tilde{V}$ is $V(\sigma) \setminus (Y_1 \cup Y_3 \cup Y_5)$.)

The second problem can be cured by thickening the boundaries slightly along the bounding rays, for example along dynamic rays at nearby angles: choose $\eta > 0$ and label the angles of the dynamic rays bounding Y_j by $\vartheta < \vartheta'$, so that $Y_j \setminus K(P_c) = \left(\bigcup_{\varphi\in(\vartheta,\vartheta')} R_c(\varphi)\right) \cap V(\sigma)$ (where $R_c(\varphi)$ is again the dynamic ray at angle φ). We define the η-thickening of Y_j as $Y_j^\eta := \left(Y_j \cup \bigcup_{\varphi\in(\vartheta-\eta,\vartheta'+\eta)} R_c(\varphi)\right) \cap V(\sigma)$. Here η has to be chosen sufficiently small so that all required dynamic rays are defined. Moreover, one needs to control the additional rays at low potentials so that they do not "unexpectedly" leave D at low potentials; this can be done by choosing η sufficiently small, or by using only the tails of the rays until they intersect D for the first time. A similar thickening can be done for all the sets Y_0, Y_j and Z_j; we denote the resulting thickened sets by Y_0^η, Y_j^η and Z_j^η, respectively.

Let

$$U := \left(Y_0^\eta \cup \bigcup_{j=1}^{q-1} (Y_{jn}^\eta \cup Z_{jn}^\eta)\right) \cap V(\sigma/2^n) \quad \text{and} \quad V := P_c^{\circ n}(U).$$

Then $P_c^{\circ n} \colon U \to V$ is indeed a quadratic-like map in the sense of Douady and Hubbard. Of course, there is a considerable amount of freedom in the choice of σ and in the two thickening steps. However, all the added points will eventually escape from U, and the (filled-in) Julia set of $P_c^{\circ n}\colon U \to V$ is independent of all the choices: this is clear for the points added in the thickening, and for the points in D this follows as soon as the orbit eventually leaves D (the only point in D for which the orbit stays in D forever is α). This implies that this Julia set is completely contained in $\tilde{U}$ (all choices even yield hybrid equivalent quadratic-like maps). Therefore we can assume that we have chosen the sets U and V for each parameter of the p/q-wake in a way that the boundary of U (and thus of V as well) depends continuously on c with respect to the Hausdorff topology on $\mathbb{C}$. We call the (filled-in) Julia set of the quadratic-like map just constructed the *little (filled-in) Julia set*. The polynomial P_c becomes n-renormalizable whenever the filled-in Julia set of this quadratic-like map is connected.

DEFINITION 3.1 (Crossed renormalization locus).
The *crossed n-renormalization locus* within the p/qn-wake, denoted $C^n_{p,q}$, is the subset of the p/qn-wake of $\mathcal{M}$ for which the little Julia set is connected.

Therefore, by definition, all P_c with $c \in C^n_{p,q}$ are n-renormalizable. In fact, this renormalization is of the type that we want:

LEMMA 3.2 (Type of renormalization).
For all $c \in C^n_{p,q}$ the polynomial P_c is immediately n-renormalizable of crossed type.

PROOF. The little Julia set is contained in $\tilde{U} \subset Y_0 \cup Y_n \cup Y_{2n} \cup \ldots \cup Y_{(q-1)n} \cup Z$, and the critical orbit of the little Julia set first visits the sectors Y_0, Y_n, Y_{2n}, $\ldots$, $Y_{(q-1)n}$. Therefore, the little Julia set meets the interiors of all the sectors Y_0, Y_n, Y_{2n}, $\ldots$, $Y_{(q-1)n}$, but of no other Y_i. The image of the little Julia set is then contained in $Y_1 \cup Y_{n+1} \cup Y_{2n+1} \cup \ldots \cup Y_{(q-1)n+1}$. Therefore, the little Julia set and its image are different. The α-fixed point is their only common point and disconnects both of them. $\square$

Obviously, $C^n_{p,q} \subset \mathcal{M}$ because for $c \notin \mathcal{M}$, the critical orbit will eventually leave the domain of the polynomial-like map, which implies that the little Julia set is disconnected; thus $C^n_{p,q} \subset \mathcal{M}_{p/qn}$.

It may not be clear at this point that for a parameter to be crossed n-renormalizable it is necessary that our particular construction yields a connected little Julia set. We will argue in Section 3.4 that this is indeed the case.

3.2. The boundary of the renormalization locus. As noted above, the closures of the loci of simple n-renormalization are finitely many homeomorphic copies of the Mandelbrot set within itself. For each of these "little Mandelbrot sets", the boundary points that disconnect it from the rest of the Mandelbrot set are one parabolic parameter and infinitely many Misiurewicz–Thurston points of (tuned) β-type; the subset of $\mathcal{M}$ that is disconnected by such a point can be chopped off by pairs of parameter rays landing at the parabolic parameter or at such Misiurewicz–Thurston points. These facts are described in [**H**, **M3**, **D**, **S5**] and have analogues for crossed renormalization. We will describe them in this section.

We will need the following folklore result several times:

LEMMA 3.3 (Correspondence between dynamic and parameter ray pairs). *Let $\langle\vartheta_1, \vartheta_2\rangle$ be a preperiodic ray pair in the dynamics of P_c with the property that $\langle\vartheta_1, \vartheta_2\rangle$ separates the critical value from the critical point and from all dynamic rays $R_c(2^k\vartheta_1)$ and $R_c(2^k\vartheta_2)$, for $k \geq 1$. Then the parameter rays $R_{\mathcal{M}}(\vartheta_1)$ and $R_{\mathcal{M}}(\vartheta_2)$*

land together at a common Misiurewicz–Thurston parameter, and this parameter ray pair separates any parameter c' from 0 *iff in the dynamics of $P_{c'}$ the dynamic ray pair $\langle\vartheta_1, \vartheta_2\rangle$ separates the critical value c' from* 0*.*

Moreover, if in the dynamics of P_c additional preperiodic dynamic rays $P_c(\vartheta_j)$ land together with the rays $P_c(\vartheta_1)$ and $P_c(\vartheta_2)$, then the parameter rays $P_{\mathcal{M}}(\vartheta_j)$ land at the same point as $P_{\mathcal{M}}(\vartheta_1)$ and $P_{\mathcal{M}}(\vartheta_2)$.

PROOF. Let $U \subset \mathbb{C}$ be the largest connected neighborhood of c with the property that every $c' \in U$ has a preperiodic dynamic ray pair $\langle\vartheta_1, \vartheta_2\rangle$ that lands at a repelling preperiodic point, and so that this ray pair separates the critical value from the critical point and from all dynamic rays $R_{c'}(2^k\vartheta_1)$ and $R_{c'}(2^k\vartheta_2)$, for $k \geq 1$. Then U is easily seen to be open (repelling periodic points remain repelling under small perturbations, and if they are landing points of periodic rays then they keep these rays under small perturbations; moreover, rays and endpoints depend analytically on the parameter, and so does the critical value). Therefore, every $c' \in (\partial U) \setminus \mathcal{M}$ is on one of the parameter rays $R_{\mathcal{M}}(2^k\vartheta_1)$ or $R_{\mathcal{M}}(2^k\vartheta_2)$ (outside of $\mathcal{M}$, all periodic orbits are repelling, so the only way for a repelling periodic point to lose a periodic ray under perturbation is when the ray fails to exist, which happens when a forward image of the ray contains the critical value). More precisely, since for $c \in U$ the ray pair $\langle\vartheta_1, \vartheta_2\rangle$ separates the critical value from all rays on the forward orbit of the ray pair, $(\partial U) \setminus \mathcal{M}$ consists exactly of the two parameter rays $R_{\mathcal{M}}(\vartheta_1)$ and $R_{\mathcal{M}}(\vartheta_2)$.

On the other hand, every $c' \in (\partial U) \cap \mathcal{M}$ is such that at least one of the dynamic rays $R_{c'}(\vartheta_1)$ or $R_{c'}(\vartheta_2)$ has a forward iterate that lands at a parabolic periodic point or at the critical value (at other parameters, the landing points of the rays $R_{c'}(2^k\vartheta_i)$ depend analytically on c', together with their rays). Both conditions describe finitely many parameters, so $\partial U \cap \mathcal{M}$ is finite.

Since $\partial U \cap \mathcal{M}$ must contain all limits points of the two parameter rays $R_{\mathcal{M}}(\vartheta_1)$ and $R_{\mathcal{M}}(\vartheta_2)$, it follows that ∂U consists of $R_{\mathcal{M}}(\vartheta_1)$ and $R_{\mathcal{M}}(\vartheta_2)$, together with a single point in $\mathcal{M}$ at which both of these parameter rays land. Since ϑ_1 and ϑ_2 are strictly preperiodic, this landing point is a Misiurewicz–Thurston parameter.

We already proved that for all $c' \in U$ there is a dynamic ray pair $\langle\vartheta_1, \vartheta_2\rangle$ that separates the critical value from the critical point and from all rays $R_{c'}(2^k\vartheta_1)$ and $R_{c'}(2^k\vartheta_2)$. It remains to show that such a ray pair exists only for $c' \in U$. It is clear that if $c' \notin \mathcal{M}$, i.e. if the parameter c' is on some parameter ray $R_{\mathcal{M}}(\vartheta)$ and hence the critical value is on the dynamic ray $R_{c'}(\vartheta)$, we must have $\vartheta \in (\vartheta_1, \vartheta_2)$, hence $c' \in U$. A perturbation argument as above shows the claim also for $c' \in \partial\mathcal{M}$ and then for $c' \in \mathcal{M}$.

If, for the dynamics of P_c, an additional dynamic ray $R_c(\vartheta_j)$ lands together with $R_c(\vartheta_1)$ and $R_c(\vartheta_2)$, this continues to be the case for all $c' \in U$ by the same perturbation reasons. This also holds at the common landing point of the parameter rays $R_{\mathcal{M}}(\vartheta_1)$ and $R_{\mathcal{M}}(\vartheta_2)$ and in some parameter neighborhood of it (at the landing point, all periodic orbits are repelling). From here it follows easily (and is well known) that $R_{\mathcal{M}}(\vartheta_j)$ lands at the same Misiurewicz–Thurston parameter as $R_{\mathcal{M}}(\vartheta_1)$ and $R_{\mathcal{M}}(\vartheta_2)$. □

REMARK 3.4. The previous lemma is a special case of the classical "Ray Correspondence Theorem" that goes back to Douady, Hubbard, and Lavaurs: *for every $c \in \mathbb{C}$, there are angle-preserving bijections between the parameter ray pairs at periodic and preperiodic angles which separate* 0 *and c, and the characteristic periodic* *and preperiodic ray pairs in the dynamic plane of c landing at repelling orbits.* The argument given here comes from Milnor [**M3**] and essentially proves the complete theorem.

The following result specifies the location of $C^n_{p,q}$ within the limb $\mathcal{M}_{p/qn}$.

LEMMA 3.5 (The loci $C^n_{p,q}$ are connected and almost compact). *Consider a crossed renormalization component $C^n_{p,q} \subset \mathcal{M}_{p/qn}$ and let $c_{p/nq}$ be the root of $\mathcal{M}_{p/nq}$. Then $C^n_{p,q} \cup \{c_{p/nq}\}$ is compact, connected, and full. Any $c \in \mathcal{M}_{p/nq} \setminus C^n_{p,q}$ is disconnected from $C^n_{p,q}$ by a pair of preperiodic parameter rays landing at a Misiurewicz–Thurston point of α-type.*

PROOF. For every parameter $c \in \mathcal{M}_{p/qn}$ (and even within the entire p/qn-wake of $\mathcal{M}$) we can construct the quadratic-like map as in Section 3.1, based on $P_c^{\circ n}$ restricted to a thickening of the set $\tilde{U}$. The parameter c is crossed n-renormalizable iff the entire critical orbit remains within $\tilde{U}$ under $P_c^{\circ n}$. Since $c \in \mathcal{M}_{p/nq}$, the α-fixed point is the landing point of qn periodic dynamic rays with combinatorial rotation number p/qn, and all points on the backward orbit of α are landing points of qn strictly preperiodic dynamic rays.

Since $c \in \mathcal{M}$, the critical orbit of $P_c^{\circ n}$ always remains within the filled-in Julia set of P_c, so it can escape from $\tilde{U}$ only through the set

$$Z' := Z \setminus \bigcup_{j=1}^{q-1} Z_{jn} \ . \tag{1}$$

We need only be concerned with the first time the critical orbit of $P_c^{\circ n}$ enters Z'. This can only happen after s iterations of $P_c^{\circ n}$ for some integer $s \geq q$. The critical value is contained in Y_1.

Let us first consider the case $s = q$. Under $qn-1$ iterations of P_c, the sector Y_1 maps homeomorphically onto $(Y_0 \cup Z)$ (except that Y_1 needs to be restricted to an appropriate equipotential: in order to simplify notation, we will in this section omit specific mention of necessary restrictions to appropriate equipotentials). There is a unique point $z_1 \in Y_1$ that maps to $-\alpha$ under $P_c^{\circ(qn-1)}$, and z_1 is the landing point of qn preperiodic dynamic rays (see Figure 3). These rays cut Y_1 into qn closed sub-sectors that intersect only a their boundaries. For $j = 1, 2, \ldots, q-1$, there is exactly one sub-sector that maps onto Z_{jn} under $P_c^{\circ(qn-1)}$, and one sub-sector maps onto its image containing Y_0 and all Y_j. These q sub-sectors are distributed evenly around z_1, and if the critical value is contained in one of them, then the critical point will survive q iterations of $P_c^{\circ n}$ in $\tilde{U}$. However, if the critical value is in one of the remaining $qn - q$ sub-sectors at z_1, then the critical point will leave $\tilde{U}$ already after q iterations of $P_c^{\circ n}$. (Finally, if the critical value equals z_1, it will never escape.)

Now we transfer this configuration into parameter space (compare again Figure 3). By Lemma 3.3 the external angles of the qn dynamic rays bounding these sub-sectors are also the external angles of qn parameter rays of the Mandelbrot set that land at a common Misiurewicz–Thurston point and cut the complex parameter plane into qn closed sectors. The sector in parameter space containing c is bounded by the parameter rays at exactly the same external angles as the sub-sector within Y_1 containing the critical value. Among the qn parameter regions, $qn-q$ do not intersect $C^n_{p,q}$, so that the renormalization locus $C^n_{p,q}$ is contained in the q-star around a Misiurewicz–Thurston point formed by the remaining q regions (together with the Misiurewicz–Thurston point itself, corresponding to the case that the critical value equals z_1). This is the first step of chopping away subsets of $\mathcal{M}_{p/qn}$ in order to approximate $C^n_{p,q}$, and it describes for which parameters the critical orbit survives the first q iterations of $P_c^{\circ n}$ within $\tilde{U}$.

If $P_c^{\circ(qn)}(0)$ is in Y_0 or some Z_{jn}, then the critical orbit survives another $q-1$ (respectively $(q-j)-1$) iterations of $P_c^{\circ n}$ within $Y_0 \cup Y_n \cup \ldots \cup Y_{(q-1)n}$. In the next iteration of $P_c^{\circ n}$, the critical orbit can again visit Y_0 or Z, and it escapes whenever it hits a "wrong" sector Z_j. The qn dynamic rays landing at $-\alpha$ can be transported back for $(q-j)n$ iteration steps of P_c into the sector Z_{jn} or Y_0 containing $P_c^{\circ(qn)}(0)$. Transporting these rays back another $qn-1$ steps, we obtain qn preperiodic dynamic rays within the sub-sector of Y_1 containing the critical value. These qn rays cut the sub-sector into qn closed sub-sub-sectors. Among these qn sub-sub-sectors, there are q of them so that if the critical value is contained in one of these, then the critical point will survive $q+(q-j)$ iteration steps of $P_c^{\circ n}$ within $\tilde{U}$; otherwise, it will not. The index j depends of course which of the sectors Z_j the critical orbit visits first. These qn new preperiodic dynamic rays have again qn counterparts in parameter space which land at a common Misiurewicz–Thurston point, and they further subdivide the parameter region containing $C^n_{p,q}$. Of the qn sub-sub-sectors around this Misiurewicz–Thurston point, only q will intersect $C^n_{p,q}$.

This argument can be repeated: in order for the critical orbit to survive one more turn within $\tilde{U}$, there is another collection of qn sub-sub-…-sectors, and only q of them may contain the critical value. We get countably many further necessary conditions which translate into a countable collection of cuts in parameter space along pairs of parameter rays at preperiodic angles. Conversely, when a parameter c is not cut off by such a parameter ray pair, then the critical orbit will remain in $\tilde{U}$ forever, and $c \in C^n_{p,q}$. The first few cuts in the dynamic plane and in parameter space are indicated in Figure 3. All the Misiurewicz–Thurston points at which the bounding parameter rays land have the property that the critical orbit terminates at the α-fixed point after finitely many iterations, so they are of α-type.

So far, we have shown that every $c \in \mathcal{M}_{p/qn} \setminus C^n_{p,q}$ is separated from $C^n_{p,q}$ by a parameter ray pair as claimed in the lemma. To conclude the proof of the lemma, let $c_{p/qn}$ be the root of $\mathcal{M}_{p/qn}$, so that $\overline{\mathcal{M}}_{p/qn} = \mathcal{M}_{p/qn} \cup \{c_{p/qn}\}$. This set is compact, connected and full. Starting with this set, every cut by a pair of parameter rays as described above leaves a compact, connected and full set, and the countable nested intersection of compact connected and full sets is compact, connected and full. □

At the root $c = c_{p/nq}$ itself, the dynamics is not renormalizable because the α-fixed point of P_c is parabolic: we have a little Julia set contained in $\tilde{U}$, but no small thickening can yield a polynomial-like map because the parabolic fixed point α attracts points along all the nq attracting directions, so every thickening will contain an invariant set outside of the little filled-in Julia set.

The situation is similar to that for simple renormalization in the immediate case of period n: any connected component of the renormalization locus is a little Mandelbrot set $\mathcal{M}'$ which is separated from the rest of its limb within the Mandelbrot set by a countable collection of parameter ray pairs landing at Misiurewicz–Thurston points of α-type. The difference is that in the simple case, the renormalization locus does not extend over such a Misiurewicz–Thurston point, while it does extend in the crossed case in $q-1$ of the $qn-1$ directions (see again Figure 3). In both cases, the root is not renormalizable. We show in the next section that the crossed n-renormalization locus $C^n_{p,q}$ within $\mathcal{M}_{p/qn}$ is homeomorphic to $\mathcal{M}_{p/q}$, while the locus of simple qn-renormalization corresponds to the tuned copy of $\mathcal{M}$ of period qn within $\mathcal{M}_{p/qn}$ under this homeomorphism.

3.3. A homeomorphism from $C^n_{p,q}$ to the p/q-limb. For every parameter in the p/qn-wake of the Mandelbrot set, we constructed a quadratic-like map and we defined $C^n_{p,q}$ to be the subset of $\mathcal{M}_{p/qn}$ where this quadratic-like map has connected Julia set, so that the corresponding polynomials are n-renormalizable. The straightening theorem supplies a well-defined map $\chi\colon C^n_{p,q} \to \mathcal{M}$. In this section, we will show that χ is a homeomorphism onto $\mathcal{M}_{p/q}$, the p/q-limb of $\mathcal{M}$.

Recall that for every $c \in \mathcal{M}_{p/qn}$, the α-fixed point of the polynomial P_c is the landing point of exactly qn dynamic rays. These rays are permuted transitively by the dynamics of the polynomial, and the combinatorial rotation number of this cyclic permutation is p/qn. Between any pair of adjacent rays, there is a part of the Julia set of P_c. Our region U contains q of these qn sectors between adjacent rays (plus a small neighborhood around α extending into all sectors). The map $P_c^{\circ n}$ permutes these sectors transitively with combinatorial rotation number $np/qn = p/q$. The straightening map χ on $C^n_{p,q}$ thus takes images within $\mathcal{M}_{p/q}$. The root of that limb cannot be in the image of χ because the α-fixed point is rationally indifferent at the root, while we started with a repelling α-fixed point and a topological conjugation can never turn a repelling periodic point into an indifferent one.

Our map χ has the following important properties:

Proposition 3.6 (The straightening map).

(1) *The straightening map χ is a homeomorphism from $C^n_{p,q}$ onto $\mathcal{M}_{p/q}$ and from $\overline{C^n_{p,q}}$ onto $\overline{\mathcal{M}_{p/q}}$.*
(2) *Its restriction to any hyperbolic component is a biholomorphic map to another hyperbolic component and extends homeomorphically to the closures. Moreover, only hyperbolic components will map to hyperbolic components.*
(3) *The map χ reduces periods of hyperbolic components exactly by a factor n. In particular, all periods of hyperbolic components in $C^n_{p,q}$ are divisible by n.*

Proof. We begin with the second statement. Within closures of hyperbolic components, there are non-repelling periodic orbits, and their multipliers are preserved under hybrid conjugations. Since hyperbolic components are well known to be conformally parametrized by the multipliers of the attracting orbits, and these parametrizations extend homeomorphically onto the closures [**DH1**, **M3**, **S2**], the straightening map inherits the same properties. (In fact, if there are any non-hyperbolic components of the interior of $\mathcal{M}$, then the straightening map will still be

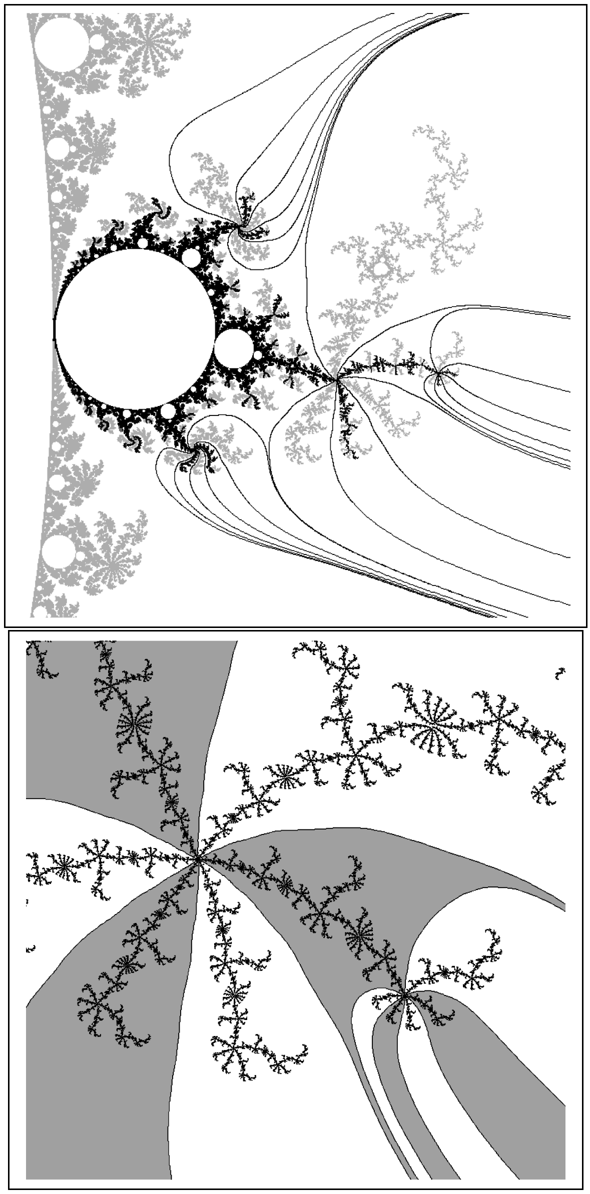


FIGURE 3. Top: the boundary of the renormalization locus $C^n_{p,q}$ in $\mathcal{M}_{p/qn}$, displayed for $n = 2$ and $p/q = 1/3$. The boundary of the entire Mandelbrot set is shown in light grey; the boundary of the crossed 2-renormalization locus is shown in dark grey, and the boundary of the simple 6-renormalization locus is shown in black. Each of these sets is a subset of the previous one. (It might help distinguish the grayscale to observe that the black locus is not disconnected by the indicated parameter ray pairs: among the five branches cut off by each group of these parameter rays, exactly two contain crossed renormalizable parameters, indicated in dark grey; the other three do not and are entirely light grey.) Bottom: some dynamic rays corresponding to parameter rays bounding the renormalization locus.

holomorphic there [**DH2**, Corollary to Proposition 13].) On the other hand, if $P_{\chi(c)}$ and thus the n-th iterate of P_c does not have an attracting periodic point, then clearly P_c itself cannot have such a point either. Thus only hyperbolic parameters can map onto hyperbolic parameters. This proves the second claim.

Since any little Julia set can meet its images (other than itself) only at repelling periodic points, it follows that the straightening map reduces periods of hyperbolic components exactly by a factor of n. This shows the third claim.

For all $c \in \mathcal{M}_{p/nq}$, the mappings $P_c^{\circ n}: U \to V = P_c^{\circ n}(U)$ as constructed in Section 3.1 are quadratic-like. By Lemma 2.3, this defines an analytic family of quadratic-like mappings (g_c); denote the straightening map by $\chi: C_{p,q}^n \to \mathcal{M}_{p/q}$. By Theorem 2.4, χ is continuous and open, and it has a local mapping degree.

Now we show that χ is surjective onto $\mathcal{M}_{p/q}$ by showing that its image is both open and closed in $\mathcal{M}_{p/q}$. Since the limb $\mathcal{M}_{p/q}$ is well known to be connected (a proof can be found in [**S3**, Corollary 5.2]) and the image is non-empty, this makes χ surjective.

The image of $C_{p,q}^n$ is open because χ is an open map. That the image is closed follows almost from continuity: let $c \in \overline{\mathcal{M}_{p/q}}$ be a boundary point of the image $\chi(C_{p,q}^n) \subset \mathcal{M}_{p/q}$. Then there is a sequence $c_1, c_2 \ldots$ in the image converging to c, and there are points $c'_1, c'_2, \ldots \in C_{p,q}^n$ with $\chi(c'_i) = c_i$. Let c' be a limit point of the sequence (c'_i) within $\mathcal{M}$. If c' is not the root of $\mathcal{M}_{p/qn}$, then $c' \in C_{p,q}^n$ because the union of $C_{p,q}^n$ with its root is compact, and $\chi(c') = c$ by continuity. However, if c' is the root, then the c_i must converge to the root of $\mathcal{M}_{p/q}$: this follows from the fact that the roots of the p/q- and p/qn-wakes of $\mathcal{M}$ have trivial fibers (this implies, but is not implied by, local connectivity of $\mathcal{M}$ at these points; for a proof, see [**H**, Theorem I] or [**S3**, Corollary 5.1]). More precisely, if the sequence (c'_i) has an infinite subsequence on the closure of the hyperbolic component of period qn in $C_{p,q}^n$, then the image sequence (c_i) will also have an infinite subsequence on the closure of the hyperbolic component of period q in $M_{p/q}$, and this sequence will converge to the root of the component. Otherwise, there must be an infinite subsequence in sublimbs of these components, and the diameters of these sublimbs tend to zero as the parameter tends to the root (the Yoccoz inequality in [**H**] makes this precise; a qualitative version of this same statement can be found in [**S3**], and either one is an essential ingredient in the proof of local connectivity of $\mathcal{M}$ at parabolic parameters). This finishes the argument that χ is surjective from $C_{p,q}^n$ to $\mathcal{M}_{p/q}$, and it extends to a continuous surjective map between the closures of both sets (which are the same sets with the roots added).

Since χ is open and has a local mapping degree over every $c \in C_{p,q}^n$, it is easy to conclude that χ is injective: $C_{p,q}^n$ contains a single hyperbolic component of period qn; it maps biholomorphically onto the unique component of period q in $\mathcal{M}_{p/q}$, and this image component has no further inverse images. Let $X \subset \mathcal{M}_{p/q}$ be the set of points with at least two preimages, counting multiplicities. Then X is open because χ is an open map, and it is closed because χ is proper (this follows from the argument given above). Since X is not all of $\mathcal{M}_{p/q}$, it is empty. Therefore, $\chi: C_{q,p}^n \to \mathcal{M}_{p/q}$ is a continuous bijection. Since the fibers of the roots of both wakes are trivial, it follows that the extension $\chi: \overline{C_{q,p}^n} \to \overline{\mathcal{M}_{p/q}}$ is also a continuous bijection and thus a homeomorphism (as a map from a compact space to a Hausdorff space). □

A natural question is whether the homeomorphism $\chi\colon C^n_{q,p} \to \mathcal{M}_{p/q}$ has nicer mapping properties than just continuity, in particular whether it extends as a quasiconformal homeomorphism to a neighborhood: for simple renormalization it is known by work of Lyubich that the straightening map from any little Mandelbrot set to $\mathcal{M}$ has such a quasiconformal extension if the little Mandelbrot set is primitive [**L**]; however, little Mandelbrot sets of satellite type do not have such quasiconformal extensions to either $\mathcal{M}$ (which is obvious) or to half the "logistical Mandelbrot set" (the parameter space of $z \mapsto \lambda z(1-z)$); they are not even quasiconformally homeomorphic to each other when the parabolic orbits at their root points have different numbers of petals [**LP**].

Proposition 3.7 (Straightening map not quasiconformal).
The straightening homeomorphism $\chi\colon \overline{C^n_{q,p}} \to \overline{\mathcal{M}_{p/q}}$ is not quasiconformal (in the sense that it does not have a quasiconformal extension to a neighborhood).

Proof. The crossed renormalization component $C^n_{p,q}$ is a subset of $\mathcal{M}_{p/qn}$, the p/qn-limb of $\mathcal{M}$. This limb contains a little Mandelbrot set $\mathcal{M}'$ of period qn containing parameters that are immediately qn-renormalizable, and this renormalization is of satellite type (in Figure 3, the boundary of this little Mandelbrot set is shown in black). In the notion of the construction described in Section 3.1, this little Mandelbrot set consists of exactly those parameters c for which the critical orbit remains in $Y_0 \cup Y_n \cup \cdots \cup Y_{(q-1)n}$; hence $\mathcal{M}' \subset C^n_{q,p}$. The restriction of the straightening map $\chi\colon \overline{C^n_{q,p}} \to \overline{\mathcal{M}_{p/q}}$ to $\mathcal{M}'$ sends $\mathcal{M}'$ to a little Mandelbrot set $\mathcal{M}'' \subset \overline{\mathcal{M}_{p/q}}$ of period q, and this restriction is compatible with the straightening maps from $\mathcal{M}'$ and $\mathcal{M}''$ coming from simple renormalization. However, this restriction does not extend as a quasiconformal homeomorphism to a neighborhood, by work of Lomonaco and Petersen [**LP**], because their numbers of parabolic petals at the root parameters are different. The claim follows. □

3.4. Our construction is complete. In the previous sections, we identified subsets $C^n_{p,q}$ of $\mathcal{M}_{p/qn}$ containing immediately n-renormalizable polynomials of crossed type. We need to show that every immediately n-renormalizable polynomial of crossed type is contained in one of the sets $C^n_{p,q}$, so that these sets and the construction above completely describe the n-renormalization locus of crossed type within the Mandelbrot set. This will be done in this section.

Proposition 3.8 (Completeness of the construction).
Any polynomial which is immediately n-renormalizable of crossed type is contained in a set $C^n_{p,q}$ for some integers p, q such that p and qn are coprime.

Proof. Let $c \in \mathcal{M}$ be such that P_c is immediately n-renormalizable of crossed type. Then the restriction of $P_c^{\circ n}$ to appropriate open simply connected domains U, V defines a quadratic-like map $P_c^{\circ n}\colon U \to V$. Let K' be its filled-in Julia set; it is connected. Since the renormalization is immediate and of crossed type, K' and $P_c(K')$ intersect exactly in α, and α separates K'. Let $q \geq 2$ be the number of connected components of $K' \setminus \{\alpha\}$.

The α-fixed point of P_c is repelling and the landing point of at least 2 periodic dynamic rays; let k be the number of these rays. By [**McM**, Theorem 7.11], we have $k \geq qn$. These rays, together with the k rays landing at $-\alpha$, cut the complex plane into $2k-1$ simply connected closed sectors Y_0, Y_1, …, Y_{k-1}, Z_1, …, Z_{k-1}, labeled similarly as in Figure 2. Since K' is connected, it contains the critical point, so K' intersects the interior of Y_0. It follows that K' intersects $\mathrm{int}(Y_{jn})$ for $j = 1, 2, \ldots, q-1$. This accounts for at least q connected components of $K' \setminus \{\alpha\}$, and since q was defined as the total number of connected components of $K' \setminus \{\alpha\}$, these are all the connected components. If k was greater than qn, then $K' \setminus \{\alpha\}$ would have more than q connected components because P_c permutes the k rays landing at α and thus the sectors between them transitively [**M3**, Lemma 2.7]. This is not the case, so $k = qn$ and thus $c \in \mathcal{M}_{p/qn}$. It follows that

$$K' \subset Y_0 \cup \bigcup_{j=1}^{q-1} Y_{jn} \cup \bigcup_{j=1}^{q-1} Z_{jn} = \tilde{U}$$

(the restriction for the Z_i follows from the symmetry of the map). Since the entire critical orbit of $P_c^{\circ n}$ is contained in K', it is contained in the sets Y_{jn} and Z_{jn} as specified above. But this means that the critical orbit never leaves the domain of the quadratic-like map as constructed in Section 3.1, so we have $c \in C^n_{p,q}$ as claimed. The set $C^n_{p,q}$ contains indeed all parameters in $\mathcal{M}_{p/qn}$ which are immediately n-renormalizable of crossed type. □

We thus have the following description of immediate renormalization of crossed type.

Corollary 3.9 (Locus of immediate crossed renormalization). *The set $C^n_{p,q}$ is exactly the subset of $\mathcal{M}_{p/qn}$ consisting of immediately n-renormalizable parameters of crossed type.*

Proof. The set $C^n_{p,q}$ is defined as a set of parameters within $\mathcal{M}_{p/qn}$ which are immediately n-renormalization of crossed type (using a particular construction), and conversely Proposition 3.8 shows that every immediately n-renormalizable parameter of crossed type within $\mathcal{M}_{p/qn}$ is indeed in $C^n_{p',q'}$ for some p', q'; since $C^n_{p',q'} \subset \mathcal{M}_{p'/q'n}$, we have $q' = q$ and thus $p' = p$. □

Remark 3.10. The limb $\mathcal{M}_{p/qn}$ does not contain any polynomial that is n-renormalizable of simple type (because simple n-renormalization is organized in the form of embedded Mandelbrot sets based at hyperbolic components of period n, each such embedded Mandelbrot set is contained in a single limb, and the p/qn-limb contains only hyperbolic components of periods qn or greater). McMullen showed that crossed n-renormalization that is not immediate is always simple m-renormalizable for some $m > 1$ strictly dividing n (see Section 4); this renormalization type cannot occur within $\mathcal{M}_{p/qn}$, either. Therefore, *$C^n_{p,q}$ is the locus of n-renormalization within $\mathcal{M}_{p/qn}$, and this renormalization is immediate and of crossed type.*

3.5. Crossed tuning. The inverse operation of simple renormalization for quadratic polynomials is known as *tuning*: take a quadratic polynomial f_0 with periodic critical point, let n be the period and let $g \in \mathcal{M}$ be arbitrary. Intuitively, one can replace all bounded Fatou components of f_0 by copies of the Julia set of g and construct a new quadratic polynomial f that is simply n-renormalizable so that the n-renormalization is hybrid equivalent to g; this new polynomial is called *f_0 tuned by g*. (To highlight the analogy to the crossed case, set $g_0(z) = z^2$: then f_0 is n-renormalizable, and the renormalization image is g_0.) This idea of Douady and Hubbard has been worked out by Haïssinsky [**Ha**]. In parameter space, there is a *tuning homeomorphism* of $\mathcal{M}$ into itself such that the origin g_0 maps to f_0 and g maps to f within the "little Mandelbrot set"; renormalization is the inverse to the tuning map, it sends the little Mandelbrot onto $\mathcal{M}$. Results towards more general constructions of tuning have been obtained by Inou [**I**].

Here, we want to propose a similar model for crossed renormalization; in view of Section 4, it will be good enough to discuss the immediate case. Let f_0 be the polynomial at the center of the unique period qn hyperbolic component in $C^n_{p,q} \subset \mathcal{M}_{p/qn}$ and g_0 the center of the period q component in $\mathcal{M}_{p/q}$ ("the rabbits with $qn-1$ respectively $q-1$ ears"). The α-fixed point disconnects the filled-in Julia set of f_0 into qn connected components, and q of them together contain a homeomorphic copy of the filled-in Julia set of g_0 (which can be extracted by a polynomial-like map); everything else is the closure of the set of points which land on this little Julia set after finitely many iterations.

There is a straightening homeomorphism $\chi\colon C^n_{p,q} \to \mathcal{M}_{p/q}$ with $\chi(f_0) = g_0$; we call its inverse the "crossed tuning map" $\tau := \chi^{-1}\colon \mathcal{M}_{p/q} \to C^n_{p,q}$. For an arbitrary $g \in \mathcal{M}_{p/q}$, set $f := \tau(g)$; we call f the polynomial f_0 *crossed-tuned by* g (we have $f_0, f \in C^n_{p,q}$ and $g_0, g \in \mathcal{M}_{p/q}$).

To describe the dynamics f, note that in the dynamics of f_0, there are n little copies of the Julia set of g_0 that all cross at the α-fixed point. When crossed-tuning f_0 by g, the n copies of the Julia set of g_0 are replaced by copies of the Julia set of g, and of course the same applies to their iterated preimages; these fill the Julia set of f densely. One difference to tuning in the simple cases is that entire filled Julia sets g_0 are replaced by filled Julia sets of g, not just individual Fatou components (of course, one may view the Fatou components of f_0 in the simple case as the filled Julia sets of $g_0(z) = z^2$; then the replaced Julia sets in the simple case are always the same and indeed most simple, while in the case of crossed renormalization, they are multi-eared rabbits.)

Remark 3.11. It is interesting to compare crossed renormalization to *intertwining surgery* by Epstein and Yampolsky [**EY**]: certain cubic polynomials can be understood as quadratic polynomials from the same limb, intertwined around their common α-fixed point. The local dynamics near the α-fixed point is similar in both cases, but globally the dynamics is quite different: in the intertwining case, there are two separate and independent critical orbits (corresponding to degree 3 of the global polynomial), while in crossed renormalization, the little Julia sets have their critical orbits related by the dynamics of P_c. Also, the preimages of the little Julia sets are distributed differently, corresponding to the different global mapping degrees.

3.6. Internal addresses. The notion of *internal addresses* has been introduced in [**S1**, **S4**] in order to efficiently describe the combinatorial structure of the Mandelbrot set. Formally, an internal address is a (finite or infinite) strictly increasing sequence of integers starting with 1, usually written as $1 \to n_1 \to n_2 \to n_3 \ldots$ with $1 < n_1 < n_2 < n_3 \ldots$. It is associated to a parameter $c \in \mathcal{M}$ as follows: all parameter rays of the Mandelbrot set at periodic angles land in pairs with equal periods, and the landing point of such a ray pair is the root of a hyperbolic component of the same period. This parameter ray pair separates the component from the origin. All these ray pairs that separate c from the origin are totally ordered. Then n_1 is the lowest period of parameter ray pairs separating c from the origin; n_2 is the lowest period of parameter ray pairs separating c from the ray pair of period n_1, and so on. (If the parameter c is exactly on such a periodic parameter ray pair, the corresponding period is still part of the internal address.) The internal address is finite iff, after finitely many steps, there no longer is such a separating periodic parameter ray pair. This happens iff the parameter c is on the closure of a hyperbolic component, and the final entry in the internal address is the period of this component; if the parameter is a bifurcation point between two hyperbolic components, then their two periods show up as the last two entries in the internal address.

A pleasant property of internal addresses is that they encode the combinatorics of the parameters they describe. For instance, it is shown in [**S4**, Proposition 4.7] that a parameter is simple n-renormalizable if and only if its internal address contains the entry n, and every subsequent entry is divisible by n. After n-renormalization, an internal address $1 \to n_1 \to n_2 \to \ldots \to n_j \to n \to k_1 n \to k_2 n \to \ldots$ turns into $1 \to k_1 \to k_2 \to \ldots$ (only entries divisible by n are taken, and their periods are divided by n). The renormalization is immediate iff the entry n follows directly after the initial 1, so the internal address has the form $1 \to n \to k_1 n \to k_2 n \to \ldots$ with $2 \le k_1 < k_2 < \ldots$.

There is a corresponding statement about crossed renormalization, which we state first for the immediate case.

Proposition 3.12 (Crossed renormalization and internal address).
Let c be a parameter in $\mathcal{M}$. Then the polynomial P_c is immediately crossed n-renormalizable if and only if its internal address is of the form

$$1 \to k_1 n \to k_2 n \to k_3 n \to \ldots$$

with $2 \le k_1 < k_2 < k_3 \ldots$ (and in particular, n does not occur in the internal address). In this case, $c \in C^n_{p,q}$ with $q = k_1$ and some p coprime to qn, and the internal address of $\chi(c)$ is $1 \to k_1 \to k_2 \to k_3 \to \ldots$.

The statement is analogous to [**S4**, Proposition 4.7], and it is possible to give a similar proof. The argument we give here is similar in spirit, even though we describe the details somewhat differently.

Proof. Let $c \in \mathcal{M}$ be a parameter that is immediately n-renormalizable of crossed type and let $1 \to n_1 \to n_2 \to \ldots$ be its internal address. Then $c \in C^n_{p,q}$ for some p, q with $q \ge 2$ and p coprime to q. We have $C^n_{p,q} \subset \mathcal{M}_{p/qn}$, so we are in the p/qn-limb of $\mathcal{M}$ and thus $n_1 = qn \ge 2n$. Since $C^n_{p,q}$ is connected, it connects c to the boundary of the main cardioid of $\mathcal{M}$. Any number n_i in the internal address corresponds to a hyperbolic component of period n_i, and the two parameter rays landing at its root disconnect $C^n_{p,q}$. The hyperbolic component of period n_i must then also be in $C^n_{p,q}$, and by Proposition 3.6, n_i is divisible by n. Since χ divides periods of hyperbolic components by n, every entry in the internal address of c is divisible by n, and every entry in the internal address of $\chi(c)$ comes from an entry in the internal address of c. Therefore, the internal address of $\chi(c)$ is as claimed.

Conversely, consider a polynomial P_c with $c \in \mathcal{M}_{p/nq}$ that is not immediately n-renormalizable of crossed type. We will show that its internal address $1 \to n_1 \to n_2 \to \ldots$ is not of the given form. Since $c \in \mathcal{M}_{p/nq}$, we have $n_1 = qn$, and we will show that some subsequent n_i is not divisible by n. The critical orbit cannot forever remain in the domain $\tilde{U} = Y_0 \cup \bigcup_{j=1}^{q-1}(Y_{jn} \cup Z_{jn})$ as constructed in Section 3.1, so (as in Lemma 3.5 and especially (1)) there must be a least number s with

$$P_c^{\circ s}(0) \in Z' = \bigcup_{j=1}^{qn-1} Z_j \setminus \bigcup_{j=1}^{q-1} Z_{jn} \ .$$

Each $z \in \tilde{U}$ has the property that $P_c^{\circ k}(z) \in \bigcup_{j=1}^{qn-1} Y_j$ for $k = 1, 2, \dots, n-1$, so s must be divisible by n. The main task in the proof is to construct a Misiurewicz–Thurston parameter $\tilde{c}$ so that the rays landing at $\tilde{c}$ separate c from $C_{p,q}^n$ and so that the internal address of $\tilde{c}$ has an entry that is not divisible by n; the same will then follow for c.

Consider the sector Z_j containing $P_c^{\circ s}(0)$ and pull it back $s-1$ iterations along the critical orbit, so that we obtain a small sector, say $\tilde{Z} \ni c$, for which $P_c^{\circ(s-1)}\colon \tilde{Z} \to Z_j$ is a homeomorphism. The point $-\alpha$ is the base point of Z_j; similarly, let $\tilde{z} \in \partial\tilde{Z}$ be the base point of $\tilde{Z}$ so that $P_c^{\circ(s-1)}(\tilde{z}) = -\alpha$. Among the qn rays landing at $\tilde{z}$, let $R_c(\vartheta_1)$ and $R_c(\vartheta_2)$ be the ones with the smallest and largest angles: these form a ray pair at $\tilde{z}$. This ray pair subdivides Y_1 into two components. Let V be the one with $\alpha \in \partial V$; then V is disjoint from $\tilde{Z} \ni c$. After exactly $s-1$ iterations, the image of the ray pair $(\vartheta_1, \vartheta_2)$ lands at $-\alpha$ and forms the common boundary of Z and Y_0.

Pull back V for qn iterations, always choosing the branch with α on the boundary. For the first iteration, this is possible because V is disjoint from $\tilde{Z} \ni c$, and for the remaining $qn-1$ iterations the domains that are pulled back are disjoint from Y_1 and thus cannot contain the critical value. Let V' be the image after qn pullback steps; we clearly have $V' \subsetneq V$. The domain V' will be bounded by four dynamic rays: two of them are the same rays bounding Y_1 and land at α, and two more are on the backwards orbit of $(\vartheta_1, \vartheta_2)$ and land at some point $\tilde{z}'$. The angles of rays in V are contained in two intervals of positive lengths; these lengths are divided by 2 in each pull-back step. Two of the bounding rays of V' thus form a ray pair that separates α from $\tilde{z}$ and from c. Repeating this process infinitely often, we obtain a sequence of ray pairs within Y_1 converging to α and to the boundary ray pair of Y_1, so that every $z \in Y_1 \setminus \{\alpha\}$ is separated from α by all but finitely many of these ray pairs.

Pulling these ray pairs back $qn-j$ times, we obtain a sequence of ray pairs in Y_{j+1} converging to the boundary, and pulling back once more choosing the branch based at $-\alpha$ rather than α, we obtain a sequence of ray pairs in Z_j converging to the boundary. Therefore, pulling back $s-1$ further iterations, we obtain a similar sequence of ray pairs within $\tilde{Z}$ separating c from $\tilde{z}$. Choose one such ray pair $(\vartheta_1', \vartheta_2')$. By construction, its landing point is on the backwards orbit of $-\alpha$. Let s' be the number of iterations it takes to land at $-\alpha$.

Now we trace the forward orbit of the ray pair $(\vartheta_1', \vartheta_2')$. After $s-1$ iterations, $(\vartheta_1', \vartheta_2')$ will be in Z_j (where j is not divisible by n, but s is), and next time it maps to Y_{j+1}. After $qn-j$ iterations, it will be in Y_1, and after this it will need a multiple of qn iterations until it maps to the ray pair $(\vartheta_1, \vartheta_2)$ landing at $\tilde{z} \in \tilde{Z}$. It then takes $s-1$ further iterations until it maps to a ray pair landing at $-\alpha$. In total, the number of iterations it takes $(\vartheta_1', \vartheta_2')$ to land at $-\alpha$ is $s' = s-1+1+qn-j+mqn+s-1 = 2s+qn-j+mqn-1$ for some $m \in \mathbb{Z}$, hence $s'+1 \equiv -j \neq 0 \pmod{m}$ (recall that s is divisible by n while j is not).

By Lemma 3.3, there is a parameter ray pair at angles $(\vartheta_1', \vartheta_2')$ that lands at a Misiurewicz–Thurston parameter $\tilde{c}$ and separates c from $C^n_{p,q}$; for this polynomial, the rays at angles ϑ_1' and ϑ_2' land at the critical value, and s' iterations later they land at $-\alpha$. If the internal address of c has the same form as those in $C^n_{p,q}$ (in particular, all entries are divisible by n), then the same must be true for $\tilde{c}$ (compare [**S4**, Proposition 5.2], "long internal addresses"). We will derive a contradiction.

In order to do this, we have to reinterpret internal addresses in terms of kneading sequences: the latter are infinite sequences over $\{\mathtt{0}, \mathtt{1}\}$, and there is a simple algorithm to convert internal addresses into kneading sequences and back: see [**S4**, Definition 3.4]. In particular, internal addresses of the form $1 \to k_1 n \to k_2 n \to k_3 n \to$ correspond to kneading sequences that have entries $\mathtt{1}$ everywhere except at position $k_1 n$ and possibly at positions $k' n$ with $k' > k_1$. If a dynamic ray, say at angle ϑ, lands at the critical value (which is the case for $P_{\tilde{c}}$ with $\vartheta = \vartheta_1'$ or $\vartheta = \vartheta_2'$), then the dynamic rays at angles $\vartheta/2$ and $(1+\vartheta)/2$ land together at the critical point and form a ray pair. The kneading sequence can be read off as the itinerary of the angles $2^k \vartheta$, for $k = 0, 1, 2, 3, \ldots$, with respect to the partition $\mathbb{S}^1 \setminus \{\vartheta/2, (1+\vartheta)/2\}$, labeled so that ϑ is the region with label $\mathtt{1}$ (so that the kneading sequence starts with a $\mathtt{1}$) [**S4**, Definition 2.4]. In particular, if the ray at angle $2^k\vartheta$ lands at $-\alpha$, then the kneading sequence has a $\mathtt{0}$ at position $k+1$. This is the case for $P_{\tilde{c}}$ and $k = s'$, so the kneading sequence of $\tilde{c}$ has an entry $\mathtt{0}$ at a position that is not divisible by n. This is a contradiction. □

REMARK 3.13 (Angled internal addresses). Internal addresses describe the combinatorics of parameters in $\mathcal{M}$ (and of the corresponding dynamics). However, in order to completely distinguish different combinatorial classes in $\mathcal{M}$, the concept has to be extended to *angled internal addresses*: every entry n in an internal address has to encode in which sublimb of the corresponding component of period n the described parameter is. This sublimb is described by its internal angle p/q. These angled internal addresses distinguish combinatorial classes (or "fibers", when taking extra care at hyperbolic components) of the Mandelbrot set completely; compare [**S3**]. An internal address of the form $1 \to k_1 \to k_2 \to k_3 \to \ldots$ is refined to an angled internal address of the form $1_{(p_0/q_0)} \to k_{1(p_1/q_1)} \to k_{2(p_2/q_2)} \to k_{3(p_3/q_3)} \to \ldots$ so that all entries k_{j+1} and beyond are contained in the p_j/q_j-sublimb of the component with period k_j with angled internal address $1_{(p_0/q_0)} \to k_{1(p_1/q_1)} \to \ldots \to k_{j\,(p_j/q_j)}$. Note in particular that the first component, of period k_1, is an immediate bifurcation from the main component, so we always have $q_0 = k_1$.

In the angle p/q, the denominator q is redundant (it can be derived combinatorially from the internal address without angles), while p can be arbitrary (coprime to q) and distinguishes various combinatorially equivalent sublimbs of $\mathcal{M}$; for details, see [**S4**, Definition 3.8 and Theorem 3.9]. Much more strongly, for every hyperbolic component, the p/q- and p'/q-sublimbs (with p and p' coprime to q) are always homeomorphic with a homeomorphism respecting periods of hyperbolic component: this was recently established in [**DS1**].

The proposition above describes how internal addresses behave under crossed renormalization. Since renormalization preserves the parametrization of hyperbolic components by multipliers and thus by internal angles, it also preserves the angles in the angled internal address: the angled internal address $1_{(p_0/q_0)} \to k_1 n_{(p_1/q_1)} \to k_2 n_{(p_2/q_2)} \ldots$, of course with $q_0 = k_1 n$, turns into $1_{(p_0/k_1)} \to k_{1(p_1/q_1)} \to k_{2(p_2/q_2)} \ldots$

(the denominator at the main cardioid is divided by n, of course interpreting p_0/k_1 modulo n, and all the other angles are unchanged.

Based on this result, there is another approach to showing that the straightening map is a homeomorphism from $C^n_{p,q}$ to $M_{p/q}$: on a combinatorial level, this follows from internal addresses. If the Mandelbrot set is locally connected (every fiber is a single point), then its topology is completely described by its combinatorics, and we get an actual homeomorphism. Without assuming local connectivity, the problem is reduced to a local one on every fiber and can easily be settled using well-known properties of the straightening map: the only fibers that possibly consist of more than one point are infinitely often simply renormalizable, and are thus covered by the domain of the continuous straightening map (continuity at these points follows from Dudko's recent "Decoration Theorem" [**D**]; compare [**DS1**]).

3.7. Puzzles and tableaux. Many topological and geometric questions about quadratic polynomials have been investigated in terms of puzzles and tableaux as developed by Branner, Hubbard and Yoccoz. This method works well for non-renormalizable polynomials and, with modifications, for polynomials that are finitely many times renormalizable. An introduction to puzzles and tableaux can be found in Hubbard [**H**] and Milnor [**M2**].

It is well known that a quadratic polynomial is simple renormalizable if and only if its critical tableau is periodic, and that the tableau does not record whether or not the polynomial is several times renormalizable. In this section, we will give a similar description for crossed renormalization.

First we briefly define the critical tableau for a quadratic polynomial P_c within any limb of the Mandelbrot set such that the forward orbit of the critical point never hits the α-fixed point. The point α is the landing point of at least two dynamic rays. Denote all these rays together with their landing point α by Γ_0 and let V_0 be the neighborhood of $K(P_c)$ bounded by an arbitrary potential. For $m \geq 1$ let $V_m := P_c^{-1}(V_{m-1})$ and $\Gamma_m := P_c^{-1}(\Gamma_{m-1})$. The closures of the connected components of $V_m \setminus \Gamma_m$ are called the *puzzle pieces of depth* m; the puzzle piece containing the critical point is called the *critical puzzle piece of depth* m. Then the *critical tableau* is a two dimensional grid in which the m-th position of the k-th column is called *critical* if the point $P_c^{\circ k}(0)$ is contained in the critical puzzle piece of depth m and *non-critical* otherwise. The definition implies that all positions in the 0-th column are critical, and all positions above any critical position are critical as well. If P_c is simple n-renormalizable, then the α-fixed point does not disconnect the little Julia set, and neither does its backwards orbit (because the little Julia sets is forward invariant under $P_c^{\circ n}$); thus the little Julia set is contained in a single puzzle piece of any depth. This implies immediately that every n-th column is entirely critical whenever P_c is simply n-renormalizable, and the converse is easily seen. In fact, if every n-th column is critical, but no other column is, then this is sufficient (as well as necessary) to imply that P_c is n-simple renormalizable and every n-th column is even critical infinitely deep).

There is a similar result on crossed renormalization.

PROPOSITION 3.14 (Crossed renormalization and tableaux).
Let c be a parameter of the Mandelbrot set with both fixed points repelling and such that the forward orbit of the critical point does not contain α. Then the polynomial P_c is immediately n-renormalizable of crossed type if and only if in its critical

tableau critical positions only occur in the 0*-th,* $2n$*-th,* $3n$*-th,* $4n$*-th, ... columns (but not in the* n*-th column).*

PROOF. Let P_c be immediately n-renormalizable of crossed type. Then by Proposition 3.8 the parameter c is contained in a limb $\mathcal{M}_{p/qn}$ ($q \geq 2$) so that we can construct a polynomial-like map $P_c^{\circ n}\colon U \to V$ as in Section 3.1. We will continue to use the notation from this construction. The qn sets $(Y_0 \cup Z_1 \cup \cdots \cup Z_{qn-1})$, Y_1, Y_2, ..., Y_{qn-1} restricted to a potential $\sigma > 0$ are the puzzle pieces of depth 0 and the sets Z_{qn-1}, ..., Z_1, Y_0, Y_1, ..., Y_{qn-1} restricted to the potential $\sigma/2$ are the puzzle pieces of depth 1. In order for the polynomial to be crossed n-renormalizable, the orbit of 0 under $P_c^{\circ n}$ has to stay within $\tilde{U} = Y_0 \cup Y_n \cup \cdots \cup Y_{(q-1)n} \cup Z_n \cup \cdots \cup Z_{(q-1)n}$. Therefore, critical positions can only occur in the 0-th, qn-th, $(q+1)n$-th, $(q+2)n$-th, ... columns and the critical tableau is as claimed.

Now assume there is an integer n such that the critical tableau contains critical positions only in the 0-th, $2n$-th, $3n$-th, ... columns. All positions in the 0-th column are critical and by the dynamics of P_c around α there is a further column where a critical position occurs. By hypothesis, its number has the form qn with $q \geq 2$. The parameter c is thus contained in a limb $\mathcal{M}_{p/qn}$ of the Mandelbrot set and we can construct a polynomial-like map $P_c^{\circ n}\colon U \to V$ as in Section 3.1. It remains to prove that all points $P_c^{\circ kn}(0)$ are contained in U. Suppose not and find a $k > 1$ such that $P_c^{\circ kn}(0)$ is not contained in U. Then there is a minimal $m \geq 1$ not divisible by n such that $P_c^{\circ(kn+m)}(0)$ is contained in $Y_0 \cup Z_1 \cup \cdots \cup Z_{qn-1}$ and thus in the critical puzzle piece of depth 0. Now we have a critical position in the $(kn+m)$-th column, and this contradicts the hypothesis. □

REMARK 3.15. If a crossed renormalization is not immediate, i.e. if it is around a periodic point, then the polynomial is also simply renormalizable and thus contained in a little Mandelbrot set (see Theorem 4.1). Since the critical tableau is constant for polynomials within a little Mandelbrot set, the critical tableau does not specify whether the corresponding polynomial is crossed renormalizable of non-immediate type. In other words, the critical tableau describes exactly the first level of (simple or crossed) renormalization.

4. Crossed renormalization: the general case

In the previous section, we described the locus of crossed renormalization for the special case that the little Julia sets cross at a fixed point (the "immediate" case of crossed renormalization). The general case is that the little Julia sets cross at a periodic point of some period $m > 1$. By the work of McMullen [**McM**] it is known that the general case can conveniently be reduced to the immediate case:

THEOREM 4.1 (The general case of crossed renormalization).
Let the polynomial P_c *be crossed* n*-renormalizable so that the crossing point of the little Julia set and its images has period* $m > 1$*. Then* P_c *is simple* m*-renormalizable and the corresponding quadratic-like map is immediately* n/m*-renormalizable of crossed type.*

Conversely, the image of any crossed m*-renormalizable polynomial under a tuning map of period* k *is crossed* m*-renormalizable of period* km*, and the period of the intersection point of the little Julia set and its images is multiplied by* k *as well.*

PROOF. Let x be a periodic point where the little Julia set crosses one of its forward images, and let $m > 1$ be its period. Obviously, m divides n, and $m < n$ because if the first return map of x was already the first return map of the little Julia set, then the little Julia set could not cross any of its forward images at x.

Since x disconnects the little Julia set and at least one of its forward images, it must be the landing point of at least four dynamic rays. The first return map of x must then permute these rays transitively (see [**M3**, Lemma 2.7] or [**S2**, Lemma 2.4]), so the ray period is a proper multiple of m.

All the dynamic rays of period m or dividing m that do not land alone partition the complex plane into some finite number of open pieces. We will consider this situation from the point of view of the m-th iterate of P_c. Then we obtain a partition formed by the fixed rays of $P_c^{\circ m}$ (using all the fixed rays, whether they land alone or not), and each piece is a "basic region" in the sense of Goldberg and Milnor [**GM**] (they exclude the case that some fixed points of $P_c^{\circ m}$ coincide; the finitely many parabolic parameters where this happens can easily be dealt with at the end, removing punctures in tuned copies of the Mandelbrot set). The little Julia set of the crossed renormalization is contained within the closure of a single basic region, and the same applies to each of its images under iteration: otherwise, the little Julia set would have to extend over a landing point of a dynamic ray pair of period m for P_c, and such a landing point itself has period at most m. At such a point, the little Julia set must meet its m-th forward image, so this point must be x. But the rays landing at x have periods greater than m.

By [**GM**, Lemma 1.6], each basic region contains exactly one fixed point of $P_c^{\circ m}$. Since crossing points of forward images of the little Julia set are fixed points of $P_c^{\circ m}$, any two forward images of the little Julia set are in different basic regions, except those which cross at a point on the forward orbit of x. Let $\tilde{V}$ be the basic region containing the critical point and thus the little Julia set. Then $\tilde{V}$ contains $P_c^{\circ(jm)}(0)$ for $j = 0, 1, 2, 3\dots$ because all these are contained in the little Julia set or those of its forward images that cross at x. All the other points on the critical orbit are contained within different images of the little Julia set and thus within different basic regions.

Let $\tilde{U}$ be the subset of $\tilde{V}$ that is mapped onto $\tilde{V}$ under $P_c^{\circ m}$. We claim that $\tilde{U}$ and $\tilde{V}$ can be thickened slightly to two regions U, V so that $P_c^{\circ m}\colon U \to V$ is a quadratic-like map with connected Julia set.

To see this, we first transport $\tilde{V}$ back m iterations of P_c along the critical orbit $0 \in \tilde{U}$, $P_c(0)$, $\dots$, $P_c^{\circ m}(0) \in \tilde{V}$. This pull-back will avoid $\tilde{V}$ except at the beginning and end, so $P_c^{\circ m}\colon \tilde{U} \to \tilde{V}$ is a degree two map. Since the partition boundary forming $\tilde{V}$ consists of fixed rays of $P_c^{\circ m}$, it is forward invariant, which implies $\tilde{U} \subset \tilde{V}$. And since all $P_c^{\circ(jm)}(0) \in \tilde{V}$ within some forward image of the little Julia set, it follows that all $P_c^{\circ(j-1)m}(0) \in \tilde{U}$, so the critical orbit of $P_c^{\circ m}$ will never leave $\tilde{U}$. It may happen that $\tilde{U}$ and $\tilde{V}$ have common boundary points, but this can be cured by a usual thickening procedure as in Section 3.1. Call the thickened regions U and V.

We then have a quadratic-like map $P_c^{\circ m}\colon U \to V$ with connected Julia set, so P_c is m-renormalizable. None of the first $m-1$ forward images of the little Julia set can meet the interior of $\tilde{V}$, so this renormalization is simple.

By construction, the little Julia set for the crossed n-renormalization is contained in $\tilde{V}$ and thus also in $\tilde{U}$ (by the same argument as above for the critical orbit).

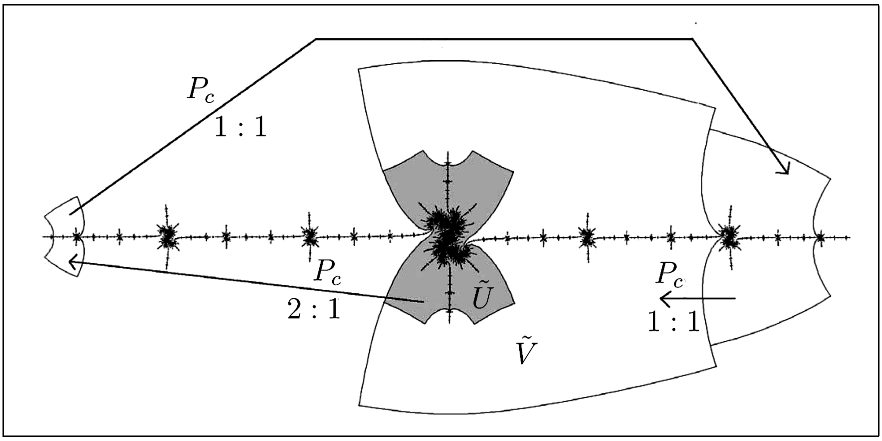


FIGURE 4. Crossed Renormalization in the general case can be reduced by a simple renormalization to an immediate crossed renormalization.

This construction preserves crossed renormalizability but reduces its period by m. Therefore, the renormalized map is still crossed n/m-renormalizable; the crossing point of this renormalization now has period one, so this crossed renormalization is immediate. This proves the first claim.

The proof of the converse statement is straightforward. □

Any crossed renormalization is thus either immediate, or it is the image of an immediate crossed renormalization under a simple renormalization. Crossed n-renormalization around a periodic point of period m can occur only if m strictly divides n; every connected component of the corresponding locus is then homeomorphic to a certain sublimb of the Mandelbrot set. This homeomorphism is a composition of a simple renormalization map of period m (a restriction of a homeomorphism sending a little Mandelbrot set of period m to $\mathcal{M}$; it reduces to the identity in the immediate case $m = 1$), followed by a homeomorphism from $C^n_{p,q} \to \mathcal{M}_{p/q}$ as constructed in Section 3.3.

All the considerations from Section 3 can now be transferred easily to the general case. Any connected component of the crossed n-renormalization locus around a periodic point of period m can be obtained from a little Mandelbrot set by chopping off subsets of $\mathcal{M}$ bounded by pairs of parameter rays at preperiodic angles (and this little Mandelbrot set itself is obtained from $\mathcal{M}$ by chopping off at further preperiodic ray pairs, together with the periodic ray pair of the root). We can also state explicitly which internal addresses allow crossed renormalizations.

COROLLARY 4.2 (Internal addresses for crossed renormalization).
A parameter of the Mandelbrot set is n-renormalizable of crossed type around a periodic point of period m if and only if m strictly divides n and its internal address is of the form

$$1 \to n_1 \to \ldots \to n_j \to m \to k_1 n \to k_2 n \to k_3 n \to \ldots$$

with $1 < k_1 < k_2 < k_3 \ldots$. *Crossed n-renormalization turns this internal address into* $1 \to k_1 \to k_2 \to k_3 \to \ldots$. *In the angled internal address, the angles at each* k_j *after renormalization are the same as those at the corresponding* $k_j n$ *before.* □

References

[BF] Bodil Branner and Núria Fagella, *Quasiconformal surgery in holomorphic dynamics*, Cambridge Studies in Advanced Mathematics, vol. 141, Cambridge University Press, Cambridge, 2014. With contributions by Xavier Buff, Shaun Bullett, Adam L. Epstein, Peter Haïssinsky, Christian Henriksen, Carsten L. Petersen, Kevin M. Pilgrim, Tan Lei and Michael Yampolsky. MR3445628

[DH0] Adrien Douady and John Hamal Hubbard, *Itération des polynômes quadratiques complexes* (French, with English summary), C. R. Acad. Sci. Paris Sér. I Math. **294** (1982), no. 3, 123–126. MR651802

[DH1] Adrien Douady and John Hamal Hubbard, *Étude dynamique des polynômes complexes. Partie I* (French), Publications Mathématiques d'Orsay [Mathematical Publications of Orsay], vol. 84, Université de Paris-Sud, Département de Mathématiques, Orsay, 1984. MR762431

[DH2] Adrien Douady and John Hamal Hubbard, *On the dynamics of polynomial-like mappings*, Ann. Sci. École Norm. Sup. (4) **18** (1985), no. 2, 287–343. MR816367

[D] Dzmitry Dudko, *The decoration theorem for Mandelbrot and multibrot sets*, Int. Math. Res. Not. IMRN **13** (2017), 3985–4028, DOI 10.1093/imrn/rnw135. MR3671509

[DS1] Dzmitry Dudko and Dierk Schleicher, *Homeomorphisms between limbs of the Mandelbrot set*, Proc. Amer. Math. Soc. **140** (2012), no. 6, 1947–1956, DOI 10.1090/S0002-9939-2011-11047-5. MR2888182

[DS2] Dzmitry Dudko and Dierk Schleicher: *Core entropy of quadratic polynomials*; with an appendix by Wolf Jung. Arnold J. Math., to appear.

[EY] Adam Epstein and Michael Yampolsky, *Geography of the cubic connectedness locus: intertwining surgery* (English, with English and French summaries), Ann. Sci. École Norm. Sup. (4) **32** (1999), no. 2, 151–185, DOI 10.1016/S0012-9593(99)80013-5. MR1681808

[GM] Lisa R. Goldberg and John Milnor, *Fixed points of polynomial maps. II. Fixed point portraits*, Ann. Sci. École Norm. Sup. (4) **26** (1993), no. 1, 51–98. MR1209913

[H] John Hamal Hubbard, *Local connectivity of Julia sets and bifurcation loci: three theorems of J.-C. Yoccoz*, Topological methods in modern mathematics (Stony Brook, NY, 1991), Publish or Perish, Houston, TX, 1993, pp. 467–511. MR1215974

[Ha] Peter Haïssinsky, *Modulation dans l'ensemble de Mandelbrot* (French, with English and French summaries), The Mandelbrot set, theme and variations, London Math. Soc. Lecture Note Ser., vol. 274, Cambridge Univ. Press, Cambridge, 2000, pp. 37–65. MR1765084

[I] Hiroyuki Inou, *Surgery construction for inverse branches of renormalization operator.* Preprint, `https://www.math.kyoto-u.ac.jp/~inou/preprint/ surgery.pdf`.

[LP] Luna Lomonaco and Carsten Lunde Petersen, *On quasi-conformal (in-) compatibility of satellite copies of the Mandelbrot set: I*, Invent. Math. **210** (2017), no. 2, 615–644, DOI 10.1007/s00222-017-0737-1. MR3714513

[L] Mikhail Lyubich, *Feigenbaum-Coullet-Tresser universality and Milnor's hairiness conjecture*, Ann. of Math. (2) **149** (1999), no. 2, 319–420, DOI 10.2307/120968. MR1689333

[M1] John Milnor, *Dynamics in one complex variable*, 3rd ed., Annals of Mathematics Studies, vol. 160, Princeton University Press, Princeton, NJ, 2006. MR2193309

[M2] John Milnor, *Local connectivity of Julia sets: expository lectures*, The Mandelbrot set, theme and variations, London Math. Soc. Lecture Note Ser., vol. 274, Cambridge Univ. Press, Cambridge, 2000, pp. 67–116. MR1765085

[M3] John Milnor, *Periodic orbits, externals rays and the Mandelbrot set: an expository account* (English, with English and French summaries), Astérisque **261** (2000), xiii, 277–333. Géométrie complexe et systèmes dynamiques (Orsay, 1995). MR1755445

[McM] Curtis T. McMullen, *Complex dynamics and renormalization*, Annals of Mathematics Studies, vol. 135, Princeton University Press, Princeton, NJ, 1994. MR1312365

[R] Johannes Riedl: *Renormierung in der Komplexen Dynamik.* In: *Nichtlineare Dynamik* (R. Mayer-Spasche, ed.), FakTUM 13, München (1997), 197–210.

[RS] Johannes Riedl and Dierk Schleicher, *On the locus of crossed renormalization*, Sūrikaisekikenkyūsho Kōkyūroku **1042** (1998), 11–31. Problems on complex dynamical systems (Japanese) (Kyoto, 1997). MR1667914

[S1] Dierk Schleicher, *Internal addresses in the Mandelbrot set and irreducibility of polynomials*, ProQuest LLC, Ann Arbor, MI, 1994. Thesis (Ph.D.)–Cornell University. MR2691195

[S2] Dierk Schleicher, *Rational parameter rays of the Mandelbrot set* (English, with English and French summaries), Astérisque **261** (2000), xiv–xv, 405–443. Géométrie complexe et systèmes dynamiques (Orsay, 1995). MR1755449

[S3] Dierk Schleicher, *On fibers and local connectivity of Mandelbrot and Multibrot sets*, Fractal geometry and applications: a jubilee of Benoît Mandelbrot. Part 1, Proc. Sympos. Pure Math., vol. 72, Amer. Math. Soc., Providence, RI, 2004, pp. 477–517. MR2112117

[S4] Dierk Schleicher, *Internal addresses of the Mandelbrot set and Galois groups of polynomials*, Arnold Math. J. **3** (2017), no. 1, 1–35, DOI 10.1007/s40598-016-0042-x. MR3646529

[S5] Dierk Schleicher, *The dynamics of complex polynomials*. Monograph, in preparation.

[T] William P. Thurston, *Entropy in dimension one*, Frontiers in complex dynamics, Princeton Math. Ser., vol. 51, Princeton Univ. Press, Princeton, NJ, 2014, pp. 339–384. MR3289916

Siemens Aktiengesellschaft, Otto-Hahn-Ring 6, D-81739 Munich, Germany
Email address: johannes.riedl@siemens.com

Aix-Marseille Université, Institut de Mathématiques de Marseille, UMR 7373, Campus de Luminy – 163 avenue de Luminy, F-13288 Marseille cedex 09, France
Email address: dierk.schleicher@gmx.de

CONM 744

Dynamics: Topology and Numbers • Moree et al., Editors

This volume contains the proceedings of the conference Dynamics: Topology and Numbers, held from July 2–6, 2018, at the Max Planck Institute for Mathematics, Bonn, Germany.

The papers cover diverse fields of mathematics with a unifying theme of relation to dynamical systems. These include arithmetic geometry, flat geometry, complex dynamics, graph theory, relations to number theory, and topological dynamics.

The volume is dedicated to the memory of Sergiy Kolyada and also contains some personal accounts of his life and mathematics.

AMS
www.ams.org

ISBN 978-1-4704-5100-4

9 781470 451004

CONM/744

AMS