Πρακτικά του Διημέρου για τον Δημήτρη Γατζούρα

Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών
Τμήμα Μαθηματικών
Αθήνα, 12-13 Μαρτίου 2022

Ευχαριστίες

Ευχαριστούμε θερμά:

- τους συγγενείς, συναδέλφους, φίλους και μαθητές του Δημήτρη που συμμετείχαν στο διήμερο.
- τους ομιλητές για τις ομιλίες τους και για τα κείμενα που συνεισέφεραν στα Πρακτικά.
- την Κοσμητεία της Σχολής Θετικών Επιστημών του Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών για την οικονομική υποστήριξη της έκδοσης των Πρακτικών αυτού του διημέρου.

Δημήτρης Γατζούρας

Ο Δημήτρης Γατζούρας γεννήθηκε το 1962. Έλαβε το πτυχίο του από το Τμήμα Μαθηματικών του ΕΚΠΑ και το διδακτορικό του δίπλωμα, με επιβλέποντα καθηγητή τον Steven Lalley, από το Πανεπιστήμιο Purdue το 1992. Στη συνέχεια είχε μεταδιδακτορικές θέσεις στο Πανεπιστήμιο Yale, το Πανεπιστήμιο Κρήτης, το Πανεπιστήμιο Cambridge και το Πανεπιστήμιο Purdue.

Επέστρεψε στην Ελλάδα, αρχικά στο Τμήμα Μαθηματικών του Πανεπιστημίου Κρήτης (πρόγραμμα Ερευνητών Εξωτερικού της ΓΓΕΤ) την περίοδο 1999-2001, και στη συνέχεια διετέλεσε μέλος ΔΕΠ του Γενικού Τμήματος του Γεωπονικού Πανεπιστημίου έως το 2016. Υπήρξε επίσης Επισκέπτης Καθηγητής στο Πανεπιστήμιο της California, San Diego. Από το 2016 υπηρετούσε ως Καθηγητής στο Τμήμα Μαθηματικών του ΕΚΠΑ.

Ο Δημήτρης Γατζούρας χαρακτηριζόταν από το εύρος των ερευνητικών ενδιαφερόντων του τα οποία εκτείνονταν σε διάφορες περιοχές της Μαθηματικής Ανάλυσης και της Θεωρίας Πιθανοτήτων. Το ερευνητικό του έργο είναι διεθνώς αναγνωρισμένο και συμπεριλαμβάνει δημοσιεύσεις πολύ υψηλής στάθμης στη Θεωρία Μέτρου, τη Θεωρία Πιθανοτήτων, την Αρμονική Ανάλυση, την Κυρτή Γεωμετρική Ανάλυση και την Εργοδική Θεωρία. Υπήρξε ιδιαίτερα αφοσιωμένος δάσκαλος: στο Τμήμα Μαθηματικών του ΕΚΠΑ δίδαξε πλήθος διαφορετικών μαθημάτων τόσο σε προπτυχιακό όσο και σε μεταπτυχιακό επίπεδο. Είχε πολύ σημαντική επιρροή στη νέα γενιά των αποφοίτων του Τμήματος και καθοδήγησε μεγάλο αριθμό μεταπτυχιακών διπλωματικών εργασιών. Ο Δημήτρης έφυγε αναπάντεχα τον Δεκέμβρη του 2020. Το κενό που αφήνει στο προπτυχιακό και μεταπτυχιακό πρόγραμμα του Τμήματος είναι πολύ μεγάλο.

Η κοινότητα του τμήματος θα τον θυμάται πάντα με συγκίνηση και θα κρατήσει ζωντανή τη μνήμη του.

Dimitris Gatzouras

Dimitris Gatzouras was born in 1962. He obtained his degree from the Department of Mathematics of NKUA and his PhD from Purdue University in 1992 (advisor: Steven Lalley). He has held postdoctoral positions at Yale University, the University of Crete, the University of Cambridge and Purdue University.

He returned to Greece, initially in the Mathematics Department of the University of Crete (program for researchers from abroad of the GSRT) from 1999 to 2001, and then was a faculty member of the General Department of the Agricultural University of Athens until 2016. He was also a Visiting Professor in the University of California at San Diego. Since 2016 he served as a Professor in the Department of Mathematics of NKUA.

Dimitris Gatzouras was characterized by the wide range of his research interests. These extended in several areas of Mathematical Analysis and Probability Theory. His research is internationally recognised and consists of publications of a very high standard, in Measure Theory, in Probability Theory, in Harmonic Analysis, in Convex Geometric Analysis and in Ergodic Theory. He was a particularly devoted Teacher: in the Department of Mathematics of NKUA he taught a large number of different courses at the undergraduate as well as the graduate level. He exerted a very important influence on the new generation of the Department's graduates and supervised a large number of postgraduate theses.

Dimitris passed away unexpectedly in December 2020. His departure has created a very wide vacuum in the undergraduate and graduate program of the Department.

Our department's community will always remember him affectionately and will keep his memory alive.

Dimitris Gatzouras
Short Curriculum Vitae

https://sites.google.com/view/dgatzouras/home

Research Areas and Interests

- Fractal Geometry, Hausdorff dimension and measure, Minkowski dimension and Minkowski content.
- Probability measures and random walks on general locally compact groups and their harmonic analysis.
- Convex Geometry, Random Polytopes.
- Ergodic Theory and Dynamical Systems.
- Probability Theory and Stochastic Processes.

Education

Diploma (B.Sc.): Mathematics, University of Athens, November 1986.
M.Sc.: Department of Statistics, Purdue University, May 1989.
Ph.D.: Department of Statistics, Purdue University, July 1992.

Ph.D. Thesis

Self-Affine Fractals: Deterministic and Random Constructions, Purdue University. (August 1992)
Thesis Advisor: S. P. Lalley.

Academic Positions

2016–20: Professor, Department of Mathematics, University of Athens, Greece.

2010–16: Associate Professor, Department of General Sciences, Agricultural University of Athens, Greece.

2002–10: Assistant Professor, Department of General Sciences, Agricultural University of Athens, Greece.

1999–01: GSRT Research Fellow, Department of Mathematics, University of Crete, Greece.

1998–99: Visiting Assistant Professor, Department of Statistics, Purdue University, U.S.A.

1995–98: Assistant Lecturer, Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, United Kingdom.

1995–98: Fellow and Director of Studies in Mathematics for 1997–98, Downing College, University of Cambridge, United Kingdom.

1993–95: Visiting Assistant Professor, Department of Mathematics, University of Crete, Greece.

1992–93: Visiting Lecturer, Department of Mathematics, Yale University, U.S.A.

Research Publications

1. Hausdorff and Box Dimensions of Certain Self-Affine Fractals (with S. Lalley), *Indiana Univ. Math. J.* **41** (1992), 533–568.

2. Statistically Self-Affine Sets: Hausdorff and Box Dimensions (with S. Lalley), *J. Theoretical Probability* **7** (1994), 437–468.

3. The Variational Principle for Hausdorff Dimension: Survey (with Y. Peres), in *Ergodic Theory of $\mathbb{Z}^d$-Actions (Warwick, 1993-1994)*, London Math. Soc. Lecture Notes, Vol. **228** (M. Pollicott, K. Schmidt eds), Cambridge University Press, 1996, pp. 113–125.

4. Invariant Measures of Full Dimension for some Expanding Maps (with Y. Peres), *Ergodic Theory and Dynamical Systems* **17** (1997), 147–167.

5. Lacunarity of Self-Similar and Stochastically Self-Similar Sets, *Trans. Amer. Math. Soc.* **352** (2000), 1953–1983.

6. On the Lattice Case of an Almost Sure Renewal Theorem for Branching Random Walks, *Advances in Applied Probability* **32** (2000), 720–737.

7. On Images of Borel Measures under Borel Mappings, *Proc. Amer. Math. Soc.* **130** (2002), 2687–2699.

8. A Spectral Radius Formula for the Fourier Transform on Compact Groups and Applications to Random Walks (with M. Anoussis), *Advances in Mathematics* **188** (2004), 425–443.

9. Lower Bound for the Maximal Number of Facets of a 0/1-Polytope (with A. Giannopoulos and N. Markoulakis), *Discrete Comp. Geom.* **34** (2005), 331–349.

10. On Summing Sequences in $\mathbb{R}^d$ (with M. Anoussis), *Illinois J. Math.* **49** (2005), 905–910.

11. A Large Deviations Approach to the Geometry of Random Polytopes (with A. Giannopoulos), *Mathematika* **53** (2006), 173–210.

12. On the Maximal Number of Facets of $0/1$ Polytopes (with A. Giannopoulos and N. Markoulakis), *GAFA Seminar Volume* 2004–2005, Lecture Notes in Mathematics, Vol. **1910** (V. D. Milman, G. Schechtman eds.), Springer, 2007, pp. 117–125.

13. On Mixing and Ergodicity in Locally Compact Motion Groups (with M. Anoussis), *J. Reine Angew. Math.* **625** (2008), 1–28.

14. Threshold for the Volume Spanned by Random Points with Independent Coordinates (with A. Giannopoulos), *Israel J. Math.* **169** (2009), 125–153.

Lecture Notes

Probability on Trees: An Introductory Climb (with Y. Peres and D. Levin), in *Lectures on Probability Theory and Statistics*, Ecole d'Eté de Probabilités de Saint-Flour XXVII-1997, (P. Bernard ed.), Lecture Notes in Mathematics, Vol. 1717, Springer Verlag, pp. 193–280.

(Y. Peres' ten lectures at the XXVII Saint-Flour Summer School in Probability, July 1997.)

Visiting Positions (long term)

- 01/2005–06/2005: Department of Mathematics, University of California at San Diego.
- 01/2008–06/2008: Department of Mathematics, University of California at San Diego.

Student Supervision

Ph.D. Students

- Georgios Zarakas: Agricultural University of Athens.
- Vasilis Sterios: University of Athens.

M.Sc. Students at the University of Athens

- Vassiliki Kouni: *An introduction to Harmonic Analysis on the line and on the circle.* (21-7-2017)
- Charis Ganotaki: *Topological dynamical systems and applications.* (25-7-2018)
- Georgios Lamprinakis: *Ergodic means on cubes.* (18-12-2018)
- Konstantinos Tsinas: *Structural ergodic theorems and applications to multiplicative functions.* (24-9-2019, co-supervised with N. Frantzikinakis and A. Giannopoulos)
- Isidoros Iakovidis: *Ergodic theoretic proof of Szemerédi's theorem.* (25-9-2019)
- Christos Nikou: *Ergodic theorems.* (9-10-2020)

- Michail Louvaris: *Random quantum states and random matrix techniques in Quantum Information Theory.*
  (19-1-2021)

- Argiro Karimali: *Unital quantum channels, majorization and Nielsen's theorem.*
  (12-3-2021, co-supervised with M. Anoussis and A. Giannopoulos)

- Vassiliki Balidou: *Fourier multipliers and Littlewood-Paley theory.*
  (1-6-2021, co-supervised with A. Giannopoulos)

- Alexandros Vlandos: *Ergodic theory of actions of locally compact groups.*
  (19-7-2021, co-supervised with A. Giannopoulos)

Seminars

(Year or semester long Seminars, met for 2hrs/week, and talks were given by the organizers.)

- *Ergodic Theory* – Crete, 1993-94.
- *Ergodic Theory* – Crete, 1994-95.
- *Ergodic Theory* – Crete, 1999-00.
- *Percolation and Disordered Systems* – Crete, 1999-00.
- *Gaussian Measures* – Crete, 2000-01.
- *Random Matrices* – Athens, 2004-05.
- *Ergodic Theory* – Athens, 2014-15.
- *Poisson Boundaries on Groups* – Athens, 2016-17.
- *Quantum Information Theory* – Athens, 2017-18.
- *Quantum Information Theory* – Athens, 2018-19.

Teaching

Undergraduate

- Purdue: Elementary Statistics*, Introduction to Probability Models, Elementary Statistical Methods.
- Yale: Calculus of functions of one variable.
- Crete: Optimization, Mathematical Statistics, Stochastic Processes.
- Cambridge: Probability and Measure*, Markov Chains.
- Agricultural University of Athens: Statistics*, Statistics–Experimentation*, Agricultural Experimentation–Statistics*, Mathematical and Statistical Software*.

- San Diego: Calculus and Analytic Geometry for Science & Engineering*, Statistical Methods*, Time Series.
- Department of Mathematics, University of Athens: Convex Analysis, Calculus II*, Real Analysis*, Measure Theory*, Calculus III, Analysis II and Applications (Department of Computer Science), Calculus I.

Graduate

- Crete: Probability Theory*.
- Cambridge: Sequential Methods*.
- Purdue: Statistical Methods.
- Agricultural University of Athens: Applied Statistics*, Exprerimental Design–Data Analysis*.
- Department of Mathematics, University of Athens: Ergodic Theory*, Harmonic Analysis, Measure Theory*.

* Repeatedly.

Participation in Research Programs and Networks

- TMR Network HARP (Harmonic Analysis and Related Problems), 2002–2006. Node in Greece: Department of Mathematics, University of Crete.
- RTN PHD (Phenomena in High Dimensions), 2004–2008 (member of the Local Steering Committee). Node in Greece: Department of Mathematics, University of Athens.
- Program EPEAEK II Pythagoras II: Asymptotic Problems in Harmonic and Convex Geometric Analysis, Department of Mathematics, University of Athens, 2005–2007 (member of research team).
- RTN CODY (Conformal Structures and Dynamics), 2007–2010. Node in Greece: Technical University of Kozani (TEI of Western Macedonia).

Contents

Spectral Radius Formulae for the Measure Algebra and the Fourier-Stieltjes Algebra

Michael Anoussis

In memory of Dimitris Gatzouras

Abstract

We present a spectral radius formula for the measure algebra of a compact group, and apply it to the study of random walks on compact groups. These are results of joint work with D. Gatzouras. We also present a spectral radius formula for the Fourier-Stieltjes algebra, obtained by P. Ohrysko and M. Roginskaya.

1 A Spectral Radius Formula for the Measure Algebra

Throughout this note G is a locally compact group, λ_G is the left Haar measure on G, and when G is compact, λ_G will be assumed to be normalized to have total mass equal to 1. We shall denote by $L^p(G)$, $1 \leq p < \infty$, the Banach space of equivalence classes of p-integrable functions on G. We will denote by $M(G)$ the space of complex, regular, Borel measures on G, and by $\|\mu\|$ the total variation norm of $\mu \in M(G)$. It is well known that when μ has a density $f \in L^1(G)$ with respect to Haar measure, then $\|\mu\| = \|f\|_1$. We will also consider the space $L^1(G)$ as a subspace of $M(G)$, identifying a function $f \in L^1(G)$ with the measure with density f. Under this identification $L^1(G)$ is an ideal of $M(G)$. We will denote by $\mathbf{1}_G$ the function identically equal to 1 on G.

If $\mathcal{A}$ is a unital Banach algebra and $x \in \mathcal{A}$, $\sigma(x)$ will denote the spectrum of x, and $\rho(x)$ the spectral radius of x. Recall that

$$\rho(x) = \lim_{n \longrightarrow \infty} \|x^n\|^{1/n} = \inf_{n \in \mathbb{N}} \|x^n\|^{1/n} .$$

The unitary dual $\widehat{G}$ of G is the set of all equivalence classes of irreducible unitary representations of G. Recall that when G is a compact group, the space $\widehat{G}$ considered as a topological space, is equipped with the discrete topology. Finally, for $R \in \widehat{G}$, d_R will stand for the dimension of R.

If $f \in L^1(G)$, f^n will denote n-fold convolution of f with itself: $f^n = f * \cdots * f$, $n \in \mathbb{N}$. Similarly, for $\mu \in M(G)$, $\mu^n = \mu * \cdots * \mu$, $n \in \mathbb{N}$, will denote n-fold convolution of μ with itself.

If $f \in L^1(G)$, $\widehat{f}$ will denote its Fourier transform:

$$\widehat{f}(R) = \int_G R\left(x^{-1}\right) f(x)\lambda_G(dx);$$

recall that $\left\|\widehat{f}(R)\right\| \leqslant \|f\|_1$ for all $R \in \widehat{G}$, and that $\widehat{(f * g)}(R) = \widehat{g}(R)\widehat{f}(R)$ for all $R \in \widehat{G}$ and $f, g \in L^1(G)$. Similarly, for $\mu \in M(G)$, $\widehat{\mu}$ will denote the Fourier transform of μ:

$$\widehat{\mu}(R) = \int_G R\left(x^{-1}\right) \mu(dx);$$

again $\|\widehat{\mu}(R)\| \leqslant \|\mu\|$ for all $R \in \widehat{G}$, and $\widehat{(\mu * \nu)}(R) = \widehat{\nu}(R)\widehat{\mu}(R)$ for all $R \in \widehat{G}$ and $\mu, \nu \in M(G)$. For a measure $\mu \in M(G)$, $\mu_{a.c.}$ and μ_s will denote its absolutely continuous and singular parts respectively, with respect to Haar measure λ_G. Recall that $\|\mu\| = \|\mu_{a.c.}\| + \|\mu_s\|$ for $\mu \in M(G)$. Finally, if a and b are non-negative numbers, $a \vee b$ denotes their maximum: $a \vee b = \max\{a, b\}$.

The following theorem which establishes a spectral formula for the measure algebra of a compact group is proved in [1].

Theorem 1.1. *Let G be a compact group, let $\widehat{G}$ be its unitary dual, and let $\mu \in M(G)$. Then*

$$\lim_{n \longrightarrow \infty} \|\mu^n\|^{1/n} = \sup_{R \in \widehat{G}} \rho\big(\widehat{\mu}(R)\big) \vee \inf_{n \in \mathbb{N}} \|(\mu^n)_s\|^{1/n}.$$

Corollary 1.2. *Let G be a compact group, let $\widehat{G}$ be its unitary dual, and let $f \in L^1(G)$. Then*

$$\lim_{n \longrightarrow \infty} \|f^n\|_1^{1/n} = \max_{R \in \widehat{G}} \rho\big(\widehat{f}(R)\big).$$

Remarks.

(1) Recall that the radical Rad $\left[L^1(G)\right]$ of the ideal $L^1(G)$ is the closed two-sided ideal in $M(G)$ consisting of those $\mu \in M(G)$ for which

$$\inf_{n \in \mathbb{N}} \|([\nu * \mu]^n)_s\|^{1/n} = 0$$

for all $\nu \in M(G)$. Thus Corollary 1.2 extends to Rad $\left[L^1(G)\right]$; i.e., for $\mu \in$ Rad $\left[L^1(G)\right]$,

$$\lim_{n \longrightarrow \infty} \|\mu^n\|^{1/n} = \sup_{R \in \widehat{G}} \rho\big(\widehat{\mu}(R)\big). \tag{1.1}$$

(2) On the other hand, (1.1) does not extend to all of $M(G)$, and this reflects the fact that $M(G)$ is asymmetric (cf. [5]). In fact (1.1) cannot hold on any asymmetric subalgebra of $M(G)$. For suppose $\mathcal{A}$ is asymmetric. Then there exists a self-adjoint measure $\mu \in \mathcal{A}$, i.e., with $\mu^* = \mu$ where $\mu^*(B) = \overline{\mu\left(B^{-1}\right)}$, whose spectrum $\sigma(\mu)$ contains a non-real complex number $\lambda = u + iv$, and we may, without loss of generality, assume that $\|\mu\| = 1$ and that $v > 0$. Then there exists a polynomial p, with $p(0) = 0$ and such that $|p(\lambda)| > \max_{x \in [-1,1]} |p(x)|$. (For example, consider the entire function $f(z) = ze^{-icz}$, where c is such that

$$|f(\lambda)| = |\lambda| e^{cv} > 1 = \max_{x \in [-1,1]} |f(x)|,$$

and approximate it by its Taylor polynomial, uniformly on the closed unit disc. We thank V. Nestoridis for this particular construction.) Since p is a polynomial without constant coefficient and $\mathcal{A}$ is an algebra, we have that $\nu = p(\mu) \in \mathcal{A}$. On the other hand, $p(\lambda) \in \sigma(\nu)$, whence $\rho(\nu) \geqslant |p(\lambda)|$. Since $\mu = \mu^*$, the operator $\widehat{\mu}(R)$ is self-adjoint and hence its spectrum is real for any $R \in \widehat{G}$; since $\widehat{\nu}(R) = p\left(\widehat{\mu}(R)\right)$, and therefore $\sigma\left(\widehat{\nu}(R)\right) = p\left(\sigma\left(\widehat{\mu}(R)\right)\right)$, we must then have that $\rho\big(\widehat{\nu}(R)\big) \leqslant \max_{x\in[-1,1]} |p(x)|$ for all $R \in \widehat{G}$. Thus the measure $\nu \in \mathcal{A}$ cannot satisfy (1.1).

(3) We would like to note that, in contrast, formula (1.1) does hold for all central measures on a compact simple Lie group. This follows from results of Ragozin [14] (Corollary 3.4 and its extension to the disconnected case, p. 228, in [14]), in conjunction with usual Gelfand theory. Note that for such groups, the algebra of central measures $Z(M(G))$ is in fact symmetric [14, p. 221].

2 An Application

If G is a compact group and μ a regular Borel probability measure on G, we shall say that the pair (G, μ) is adapted if μ is not supported by a proper closed subgroup of G. When (G, μ) is adapted, we shall say that μ is strictly aperiodic if it is not concentrated on a coset of a proper, closed, normal subgroup of G.

As an application of the spectral radius formula of Theorem 1.1 the following is proved in [1].

Theorem 2.1. *Suppose G is a compact group, and μ a regular Borel probability measure on G. Then*

$$\|\mu^n - \lambda_G\|^{1/n} \longrightarrow \sup_{R\in\widehat{G}\smallsetminus\{\mathbf{1}_G\}} \rho\big(\widehat{\mu}(R)\big) \vee \inf_{n\in\mathbb{N}} \|(\mu^n)_s\|^{1/n}.$$

In particular, $\|\mu^n - \lambda_G\| \longrightarrow 0$ iff

(1) *(G, μ) is adapted and μ is strictly aperiodic, and*

(2) *$\mu, \mu^2, \ldots$ are not all singular with respect to Haar measure λ_G.*

Moreover, the convergence $\|\mu^n - \lambda_G\| \longrightarrow 0$ takes place exponentially fast when it holds.

Corollary 2.2. *Suppose G is a compact group, and μ a regular Borel probability measure on G. Suppose further that μ is absolutely continuous with respect to Haar measure λ_G on G. Then*

$$\lim_{n\longrightarrow\infty} \|\mu^n - \lambda_G\|^{1/n} = \max_{R\in\widehat{G}\smallsetminus\{\mathbf{1}_G\}} \rho\big(\widehat{\mu}(R)\big).$$

If in particular (G, μ) is adapted and μ strictly aperiodic, then $\|\mu^n - \lambda_G\| \longrightarrow 0$ exponentially fast.

Remarks.

(1) Corollary 2.2 already appears in [10, Theorem 13 and §5.2] (see also [9]). However, Corollary 2.2 also gives the rate of decay of $\|\mu^n - \lambda_G\|$; in particular, the estimate

$$\lim_{n\longrightarrow\infty} \|\mu^n - \lambda_G\|^{1/n} = \max_{R\in\widehat{G}\smallsetminus\{\mathbf{1}_G\}} \rho\big(\widehat{\mu}(R)\big) = a$$

is sharp, in the sense that, for each $\epsilon > 0$, one has that

$$a^n \leq \|\mu^n - \lambda_G\| \leq (a+\epsilon)^n,$$

for sufficiently large n.

(2) This estimate, valid when μ has an L^1-density, is also to be compared to a well known estimate of Diaconis and Shahshahani [6], of use only when μ^n has an L^2-density: $\|\mu^n - \lambda_G\|^2 \leqslant \sum_{R\neq \mathbf{1}_G} d_R \, \|\widehat{\mu}(R)^n\|_2^2$ (here $\| \ \|_2$ denotes the Hilbert-Schmidt norm).

(3) A weak form of Theorem 2.1 seems to first have appeared in Bhattacharya [4, Theorems 2 and 3]. In the general form presented here, it also appears in Mindlin [11, Theorem 1], and in Ross and Xu [15, Theorem 4.1] (except again for the precise rate of decay to 0).

The following lemma is implicit in the proof of Theorem 3.3.5 in [16].

Lemma 2.3. *Let G be a compact group, and μ a regular Borel probability measure on G not supported by a proper closed subgroup of G. If $\rho\big(\widehat{\mu}(R)\big) = 1$ for some $R \in \widehat{G}$ with $R \neq \mathbf{1}_G$, then there exists a closed, normal, proper subgroup H of G with $\mu(gH) = 1$ for some $g \in G$.*

Proof of Theorem 2.1. We provide a proof in case μ has a non-trivial absolutely continuous part. It is straightforward to verify that $\mu^n - \lambda_G = (\mu - \lambda_G)^n$, so that

$$\begin{aligned}\|\mu^n - \lambda_G\|^{1/n} \quad &\longrightarrow \quad \sup_{R\in\widehat{G}} \rho\left(\widehat{(\mu-\lambda_G)}(R)\right) \vee \inf_{n\in\mathbb{N}} \|(\mu^n)_s\|^{1/n} \\ &= \quad \sup_{R\in\widehat{G}\setminus\{\mathbf{1}_G\}} \rho\big(\widehat{\mu}(R)\big) \vee \inf_{n\in\mathbb{N}} \|(\mu^n)_s\|^{1/n},\end{aligned}$$

by Theorem 1.1. Since $\mu_{a.c.} \neq 0$, we must have that $\|\mu_s\| < 1$. Since (G,μ) is adapted and μ is strictly aperiodic, we also have that $\rho\big(\widehat{\mu}(R)\big) < 1$ for all $R \in \widehat{G}$ with $R \neq \mathbf{1}_G$, by Lemma 2.3. It then follows from the Riemann-Lebesgue lemma that $\sup_{R\in\widehat{G}\setminus\{\mathbf{1}_G\}} \rho\big(\widehat{\mu}(R)\big) < 1$; for if $\widehat{F} \subseteq \widehat{G}$ is a finite set for which $\|\widehat{\mu}_{a.c.}(R)\| < \frac{1}{2}\,(1 - \|\mu_s\|)$ for all $R \in \widehat{G} \setminus \widehat{F}$, then

$$\|\widehat{\mu}(R)\| \leqslant \|\widehat{\mu}_{a.c.}(R)\| + \|\widehat{\mu}_s(R)\| \leqslant \tfrac{1}{2}\big(1 - \|\mu_s\|\,\big) + \|\mu_s\| = \tfrac{1}{2}\big(1 + \|\mu_s\|\,\big)$$

for all $R \in \widehat{G} \setminus \widehat{F}$, whence

$$\begin{aligned}\sup_{R\in\widehat{G}\setminus\{\mathbf{1}_G\}} \rho\big(\widehat{\mu}(R)\big) \quad &\leqslant \quad \max_{R\in\widehat{F}\setminus\{\mathbf{1}_G\}} \rho\big(\widehat{\mu}(R)\big) \vee \sup_{R\in\widehat{G}\setminus\widehat{F}} \|\widehat{\mu}(R)\| \\ &\leqslant \quad \max_{R\in\widehat{F}\setminus\{\mathbf{1}_G\}} \rho\big(\widehat{\mu}(R)\big) \vee \big(\tfrac{1}{2} + \tfrac{1}{2}\,\|\mu_s\|\big) \quad < \quad 1.\end{aligned}$$

On the other hand, $\|(\mu^n)_s\| \leq \|(\mu_s)^n\|$, which implies $\|(\mu^n)_s\|^{1/n} \leq \|\mu_s\|$.

Thus

$$\sup_{R\in\widehat{G}\setminus\{\mathbf{1}_G\}} \rho\big(\widehat{\mu}(R)\big) \vee \inf_{n\in\mathbb{N}} \|(\mu^n)_s\|^{1/n} < 1,$$

and so $\|\mu^n - \lambda_G\| \longrightarrow 0$, exponentially fast.

We prove the necessity of conditions (1) and (2). If (G, μ) is not adapted, then μ^n is concentrated on a closed proper subgroup of G. If μ is not strictly aperiodic, there is a proper, closed, normal subgroup N of G and $g \in G$ such that μ is concentrated on gN and consequently μ^n is concentrated on $g^n N$. In either case, μ^n is singular with respect to λ_G for all $n \in \mathbb{N}$. If μ^n is singular with respect to λ_G for all $n \in \mathbb{N}$, then $\|\mu^n - \lambda_G\| = 2$ for all $n \in \mathbb{N}$, whence μ^n cannot converge to λ_G in norm.

□

3 A spectral radius formula for the Fourier-Stieltjes algebra

The Fourier algebra and the Fourier-Stieltjes algebra of a locally compact group were introduced by Eymard in [7]. They are important objects of study in the area of Noncommutative Harmonic Analysis. We refer the reader to the monograph [8] for information about these algebras.

Let G be a locally compact group and $B(G)$ be the space of all functions of the form

$$\langle R(x)\xi, \eta\rangle$$

where R is a unitary representation of G in a Hilbert space H and ξ, η are vectors in H. If $f \in L^1(G)$ we denote by

$$\|f\|_{C^*(G)}$$

the C^*-norm of f, defined by

$$\|f\|_{C^*(G)} = \sup\{\|R(f)\| : R \in \widehat{G}\}.$$

The map

$$u \mapsto \sup\left\{\left|\int_G u(x)f(x)dx\right| : f \in L^1(G), \|f\|_{C^*(G)} \leq 1\right\}$$

is a norm on $B(G)$. The space $B(G)$ with this norm and pointwise multiplication is a Banach algebra, called the Fourier-Stieltjes algebra of the group G.

The Fourier algebra $A(G)$ of a locally compact group G is the space of all functions of the form

$$\langle\lambda(x)\xi, \eta\rangle$$

where λ is the left regular representation of G and ξ, η are in $L^2(G)$. It is a closed ideal of $B(G)$.

If G is abelian, the dual $\widehat{G}$ of G is a locally compact group and then the Fourier-Stieltjes transform $\mu \mapsto \widehat{\mu}$ is an isometric isomorphism from the measure algebra $M(G)$ onto the Fourier-Stieltjes algebra $B(\widehat{G})$ of the group $\widehat{G}$. Similarly, the Fourier transform $f \mapsto \widehat{f}$ is an isometric isomorphism from $L^1(G)$ onto the Fourier algebra $A(\widehat{G})$ of the group $\widehat{G}$.

Recall that the measure algebra $M(G)$ decomposes as

$$M(G) = L^1(G) \oplus M_s(G)$$

where $L^1(G)$ is identified via the Radon-Nikodym Theorem with the space of absolutely continuous measures with respect to the Haar measure and $M_s(G)$ is the space of singular measures (i.e. the space of measures supported on a set of zero Haar measure).

Arsac has proved in [3] that there exists a subspace $B_s(G)$ of the Fourier-Stieltjes algebra such that $B(G)$ decomposes as

$$B(G) = A(G) \oplus B_s(G).$$

If G is abelian, then the space of Fourier-Stieltjes transforms of measures in $M_s(G)$ is the space $B_s(\widehat{G})$ For more information about this decomposition of the Fourier-Stieljes algebra the reader is referred to [13].

Let $f \in B(G)$. We will write

$$f = f_{a.c.} + f_s$$

with $f_{a.c.} \in A(G)$ and $f_s \in B_s(G)$.

The following result was proved by Ohrysko and Roginskaya in [12].

Theorem 3.1. *Let G be a locally compact group and $f \in B(G)$. Then, the spectral radius $\rho(f)$ of f is given by the following formula*

$$\rho(f) = \|f\|_\infty \vee \inf \|(f^n)_s\|^{1/n}.$$

Remarks.

(1) We recall that if G is a locally compact group, then G is an abelian compact group if and only if its dual group $\widehat{G}$ is an abelian discrete group. Thus, it follows from Theorem 1.1 that the conclusion of Theorem 3.1 holds for discrete abelian groups and similarly it follows from Theorem 3.1 that the conclusion of Theorem 1.1 holds for abelian compact groups.

(2) A consequence of Theorem 3.1 is that the spectral radius formula of Theorem 1.1 is valid for all abelian locally compact groups.

(3) It is proved in [2] that the spectral radius formula of Theorem 1.1 is valid for locally compact motion groups.

References

[1] M. Anoussis, D. Gatzouras, *A spectral radius formula for the Fourier transform on compact groups and applications to random walks*, Adv. Math. **188** (2004), no. 2, 425–443. Addendum: Adv. Math. **202** (2006), no. 1, 287–288.

[2] M. Anoussis, D. Gatzouras, *On mixing and ergodicity in locally compact motion groups*, J. Reine Angew. Math. **625** (2008), 1–28.

[3] G. Arsac, *Sur l'espace de Banach engendré par les coefficients d'une représentation unitaire*, Publ. Dép. Math. (Lyon) **13** (1976), no. 2, 1–101.

[4] R. N. Bhattacharya, *Speed of Convergence of the n-Fold Convolution of a Probability Measure on a Compact Group*, Z. Wahrsch. Verw. Gebiete **25** (1972), 1–10.

[5] G. Brown, C. Karanikas, J. H. Williamson, *The asymmetry of $M_0(G)$*, Math. Proc. Cambridge Philos. Soc. **91** (1982), 407–433.

[6] P. Diaconis, M. Shahshahani, *Generating a random permutation with random transpositions*, Z. Wahrsch. Verw. Gebiete **57** (1981), 159–179.

[7] P. Eymard, *L'algèbre de Fourier d'un groupe localement compact*, Bull. Soc. Math. France **92** (1964), 181–236.

[8] E. Kaniuth, A. T. Lau, *Fourier and Fourier-Stieltjes algebras on locally compact groups*, Mathematical Surveys and Monographs, **231** American Mathematical Society, Providence, RI, 2018.

[9] Y. Kawada, K. Ito, *On the probability distributions on a compact group*, Proc. Phys.-Math. Soc. Japan **22** (1940), 977–998.

[10] B. M. Kloss, *Probability distributions on bicompact topological groups*, Th. Probab. Appl. **4** (1959), 237–270.

[11] D. S. Mindlin, *Rate of convergence of convolutions of random measures on a compact group*, Th. Probab. Appl. **35** (1990), 368–373.

[12] P. Ohrysko, M. Roginskaya, *The spectral radius formula for Fourier-Stieltjes algebras*, Canad. Math. Bull. **63** (2020), no. 2, 269–275.

[13] P. Ohrysko, M. Wasilewski, *Spectral theory of Fourier-Stieltjes algebras*, J. Math. Anal. Appl. **473** (2019), no. 1, 174–200.

[14] D. Ragozin, *Central Measures on Compact Simple Lie Groups*, J. Funct. Anal. **10** (1972), 212–229.

[15] K. A. Ross, D. Xu, *Norm convergence of random walks on compact hypergroups*, Math. Z. **214** (1993), 415–423.

[16] K. Stromberg, *Probabilities on a compact group*, Trans. Amer. Math. Soc. **94** (1960), 295–309.

Michael Anoussis: Department of Mathematics, University of the Aegean, 832 00 Karlovasi – Samos, Greece.

E-mail: mano@aegean.gr

Half-space depth and threshold for the measure of random polytopes

*Silouanos Brazitikos**

Dedicated to the memory of Dimitris Gatzouras

Abstract

Given a probability measure μ on $\mathbb{R}^n$, Tukey's half-space depth is defined for any $x \in \mathbb{R}^n$ by $\varphi_\mu(x) = \inf\{\mu(H) : H \in \mathcal{H}(x)\}$, where $\mathcal{H}(x)$ is the set of all half-spaces H of $\mathbb{R}^n$ containing x. We show that if μ is log-concave then

$$e^{-c_1 n} \leqslant \int_{\mathbb{R}^n} \varphi_\mu(x)\, d\mu(x) \leqslant e^{-c_2 n/L_\mu^2}$$

where L_μ is the isotropic constant of μ and $c_1, c_2 > 0$ are absolute constants. Tukey's half-space depth plays an important role in the study of the question if there exists a threshold for the expected measure of the random polytope $K_N = \mathrm{conv}\{X_1, \ldots, X_N\}$, where $X_1, X_2, \ldots$ are independent random points in $\mathbb{R}^n$ distributed according to μ. We present a general approach to this problem, which is based on the Cramer transform Λ_μ^* of μ.

1 Introduction

Our starting point is the last (chronologically) published paper of Dimitris. In a joint work with A. Giannopoulos [13] they studied the question to obtain a threshold for the expected volume of a random polytope defined as the convex hull of independent random points with a given distribution. To get a feeling of the problem, consider first the example of the hypercube $[0,1]^n$ that was first examined in [11]. This polytope has 2^n vertices and volume 1. Let $N = N(n)$, and let $Z_1, Z_2, \ldots Z_N$ be independent random variables, each uniformly distributed over $[0,1]^n$. Form the convex hull K_N of these random points, and let $V_{n,N}$ be its expected volume, that is $V_{n,N} = \mathbb{E}|K_N|$. How large should $N(n)$ be to pick up significant volume? The answer is surprisingly small. Namely, let $\kappa = 2\pi/e^{\gamma+1/2}$, where γ is Euler's constant and $\varepsilon > 0$. Then, we have that

$$\lim_{n\to\infty} \sup\{V_{n,N} \colon N \leqslant (\kappa - \epsilon)^n\} = 0$$

and

$$\lim_{n\to\infty} \inf\{V_{n,N} \colon N \geqslant (\kappa + \epsilon)^n\} = 1.$$

*The talk is based on joint works with A. Giannopoulos and M. Pafis, supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the "First Call for H.F.R.I. Research Projects to support Faculty members and Researchers and the procurement of high-cost research equipment grant" (Project Number: 1849) and the "Third Call for H.F.R.I. PhD Fellowships" (Fellowship Number: 5779).

In [13], a very general result of this type was proved. In this work, Z_i have independent identically distributed coordinates, according to a measure μ (they imposed on it some very mild assumptions) supported on a bounded interval. Using sophisticated techniques from large deviations theory, they proved that in that case there is also a sharp threshold, given by $\kappa = \mathbb{E}_\mu(\Lambda^*)$, where Λ^* is the Cramer transform of the measure μ, i.e. the Legendre transform of the log-Laplace transform of μ (see Section 3). According to a general geometric lemma, to do this, the main task is to study Tukey's half-space depth φ_μ which is defined for any $x \in \mathbb{R}^n$ by $\varphi_\mu(x) = \inf\{\mu(H) : H \in \mathcal{H}(x)\}$, where $\mathcal{H}(x)$ is the set of all half-spaces H of $\mathbb{R}^n$ containing x. In recent works with A. Giannopoulos and M. Pafis we revisit these questions.

2 Tukey's half-space depth

The first work in statistics where some form of the half-space depth appears is an article of Hodges [14] from 1955. Tukey introduced the half-space depth for data sets in [20] as a tool that enables efficient visualization of random samples in the plane. The term "depth" also comes from Tukey's article. A formal definition of the half-space depth as a way to distinguish points that fit the overall pattern of a multivariable probability distribution and to obtain an efficient description, visualization, and nonparametric statistical inference for multivariable data, was given by Donoho and Gasko in [10] (see also [19]). We refer the reader to the survey article of Nagy, Schutt and Werner [16] for an overview of this topic, with an emphasis on its connections with convex geometry, and many references. In [4] we study the expectation

$$\mathbb{E}_\mu(\varphi_\mu) := \int_{\mathbb{R}^n} \varphi_\mu(x)\, d\mu(x)$$

of φ_μ with respect to μ. The following question was asked in [1]: Does there exist an absolute constant $c \in (0,1)$ such that $\mathbb{E}_\mu(\varphi_\mu) \leqslant c^n$ for all $n \geqslant 1$ and all log-concave probability measures μ in $\mathbb{R}^n$?

We provide an affirmative answer.

Theorem 2.1. *Let μ be a log-concave probability measure on $\mathbb{R}^n$, $n \geqslant n_0$. Then, $\mathbb{E}_\mu(\varphi_\mu) \leqslant \exp\left(-cn/L_\mu^2\right)$ where L_μ is the isotropic constant of μ and $c > 0$, $n_0 \in \mathbb{N}$ are absolute constants.*

Since the quantity $\mathbb{E}_\mu(\varphi_\mu)$ is affinely invariant, we may assume that μ is isotropic. If μ is a log-concave measure on $\mathbb{R}^n$ with density f_μ, the isotropic constant of μ is defined by

$$L_\mu := \left(\frac{\sup_{x\in\mathbb{R}^n} f_\mu(x)}{\int_{\mathbb{R}^n} f_\mu(x)dx}\right)^{\frac{1}{n}} [\det \mathrm{Cov}(\mu)]^{\frac{1}{2n}},$$

where $\mathrm{Cov}(\mu)$ is the covariance matrix of μ with entries

$$\mathrm{Cov}(\mu)_{ij} := \frac{\int_{\mathbb{R}^n} x_i x_j f_\mu(x)\, dx}{\int_{\mathbb{R}^n} f_\mu(x)\, dx} - \frac{\int_{\mathbb{R}^n} x_i f_\mu(x)\, dx}{\int_{\mathbb{R}^n} f_\mu(x)\, dx}\frac{\int_{\mathbb{R}^n} x_j f_\mu(x)\, dx}{\int_{\mathbb{R}^n} f_\mu(x)\, dx}.$$

We say that a log-concave probability measure μ on $\mathbb{R}^n$ is isotropic if is centered (it has its barycenter at the origin) and $\mathrm{Cov}(\mu) = I_n$, where I_n is the identity $n \times n$ matrix. Note that a convex body K of volume 1 is isotropic if and only if the log-concave probability measure with density $L_K^n \mathbf{1}_{K/L_K}$ is isotropic. The hyperplane conjecture asks if there exists an absolute constant $C > 0$ such that

$$L_n := \max\{L_\mu : \mu \text{ is an isotropic log-concave probability measure on } \mathbb{R}^n\} \leqslant C$$

for all $n \geqslant 1$. Let us mention here that the best known upper bound for L_μ is due to Klartag and Lehec [15] (after breakthrough work of Y. Chen [8]) who showed that $L_n \leqslant c(\ln n)^4$, where $c > 0$ is an absolute constant.

In fact, Theorem 2.1 is a special case of a more general result.

Theorem 2.2. *Let μ and ν be two log-concave probability measures on $\mathbb{R}^n$, $n \geqslant n_0$, with the same barycenter. Then,*

$$\mathbb{E}_\nu(\varphi_\mu) := \int_{\mathbb{R}^n} \varphi_\mu(x)\, d\nu(x) \leqslant \exp\left(-cn/L_\nu^2\right),$$

for some absolute constants $c > 0$ and $n_0 \in \mathbb{N}$.

We also show that, apart from the value of the isotropic constant L_μ, the exponential estimate provided by Theorem 2.1 is sharp.

Theorem 2.3. *Let μ be a log-concave probability measure on $\mathbb{R}^n$. Then,*

$$\int_{\mathbb{R}^n} \varphi_\mu(x) d\mu(x) \geqslant e^{-cn},$$

where $c > 0$ is an absolute constant.

The proof of Theorem 2.2 and Theorem 2.3 makes use of several facts about isotropic log-concave probability measures. In particular, we exploit the properties of the family of the L_t-centroid bodies $Z_t(\mu)$ of μ; for any $t \geqslant 1$ the body $Z_t(\mu)$ is the centrally symmetric convex body whose support function is

$$h_{Z_t(\mu)}(y) := \left(\int_{\mathbb{R}^n} |\langle x, y\rangle|^t\, d\mu(x)\right)^{1/t}.$$

Several variants of the threshold problem for the expected volume of random polytopes have been studied. Besides [11] and [13], the articles [18] and [2], [3] address the same question for a number of cases where X_i have rotationally invariant densities. Exponential in the dimension upper and lower thresholds are obtained in [12] for the case where X_i are uniformly distributed in a simplex.

An upper threshold was obtained recently by Chakraborti, Tkocz and Vritsiou in [7] for a large family of distributions: If μ is an even log-concave probability measure supported on a convex body K in $\mathbb{R}^n$ and if $X_1, X_2, \ldots$ are independent random points distributed according to μ then for any $n < N \leqslant \exp(c_1 n/L_\mu^2)$ we have that

$$\frac{\mathbb{E}_{\mu^N}(|K_N|)}{|K|} \leqslant \exp\left(-c_2 n/L_\mu^2\right),$$

where $c_1, c_2 > 0$ are absolute constants. In [4] we obtain an analogous estimate in a more general setting.

Theorem 2.4. *Let μ be a centered log-concave probability measure on $\mathbb{R}^n$. Let $X_1, X_2, \dots$ be independent random points in $\mathbb{R}^n$ distributed according to μ and for any $N > n$ consider the random polytope $K_N = \mathrm{conv}\{X_1, \dots, X_N\}$. Then, for any centered log-concave probability measure ν on $\mathbb{R}^n$ and any $N \leqslant \exp(c_1 n/L_\nu^2)$ we have that*

$$\mathbb{E}_{\mu^N}(\nu(K_N)) \leqslant \exp\left(-c_2 n/L_\nu^2\right),$$

where $c_1, c_2 > 0$ are absolute constants.

Theorem 2.4 shows that if $N_1(n) = \exp(cn/L_n^2)$, where $c > 0$ is an absolute constant, then

$$\sup_{\mu,\nu}\Big(\sup\{\mathbb{E}\left[\nu(K_N)\right] : N \leqslant N_1(n)\}\Big) \longrightarrow 0$$

as $n \to \infty$, where the first supremum is over all centered log-concave probability measures μ and ν on $\mathbb{R}^n$.

A lower threshold is also established in [7] for the case where μ is an even κ-concave measure on $\mathbb{R}^n$ with $0 < \kappa < 1/n$, supported on a convex body K in $\mathbb{R}^n$. If $X_1, X_2, \dots$ are independent random points in $\mathbb{R}^n$ distributed according to μ and $K_N = \mathrm{conv}\{X_1, \dots, X_N\}$ as before, then for any $M \geqslant C$ and any $N \geqslant \exp\left(\frac{1}{\kappa}(\log n + 2\log M)\right)$ we have that

$$\frac{\mathbb{E}_{\mu^N}(|K_N|)}{|K|} \geqslant 1 - \frac{1}{M},$$

where $C > 0$ is an absolute constant. Since the family of log-concave probability measures corresponds to the case $\kappa = 0$, it is natural to ask for an analogue of this result for 0-concave, i.e. log-concave, probability measures. We obtain the next lower threshold for the case $\nu = \mu$.

Theorem 2.5. *Let $\delta \in (1/n^2, 1)$. Then,*

$$\inf_{\mu}\,\inf\Big\{\mathbb{E}\left[\mu((1+\delta)K_N)\right] : N \geqslant \exp\left(C\delta^{-1}\ln(2/\delta)\, n\ln n\right)\Big\} \longrightarrow 1$$

as $n \to \infty$, where the first infimum is over all log-concave probability measures μ on $\mathbb{R}^n$ and $C > 0$ is an absolute constant.

It should be noted that an exponential in the dimension lower threshold is not possible in full generality. For example, in the case where X_i are uniformly distributed in the Euclidean ball one needs $N \geqslant \exp(cn\ln n)$ points so that the volume of a random K_N will be significantly large. Thus, apart from the constants depending on δ, Theorem 2.5 is sharp. However, it provides a weak threshold in the sense that we estimate the expectation $\mathbb{E}_{\mu^N}\left(\mu(1+\delta)K_N\right)$ (for an arbitrarily small but positive value of δ) while the original question is about $\mathbb{E}_{\mu^N}\left(\mu(K_N)\right)$. We are able to "remove the δ-term", however the dependence on n is worse. More precisely, we show that there exists an absolute constant $C > 0$ such that

$$\inf_{\mu}\,\inf\Big\{\mathbb{E}\left[\mu(K_N)\right] : N \geqslant \exp(C(n\ln n)^2 u(n))\Big\} \longrightarrow 1$$

as $n \to \infty$, where the first infimum is over all log-concave probability measures μ on $\mathbb{R}^n$ and $u(n)$ is any function with $u(n) \to \infty$ as $n \to \infty$.

3 The threshold problem

In [5] we study the question to obtain a threshold for the expected measure of a random polytope defined as the convex hull of independent random points with a log-concave distribution. The general formulation of the problem is the following. Given a log-concave probability measure μ on $\mathbb{R}^n$, let $X_1, X_2, \ldots$ be independent random points in $\mathbb{R}^n$ distributed according to μ and for any $N > n$ define the random polytope

$$K_N = \mathrm{conv}\{X_1, \ldots, X_N\}.$$

Then, consider the expectation $\mathbb{E}_{\mu^N}[\mu(K_N)]$ of the measure of K_N, where $\mu^N = \mu\otimes\cdots\otimes\mu$ (N times). This is an affinely invariant quantity, so we may assume that μ is centered, i.e. the barycenter of μ is at the origin.

Given $\delta \in (0,1)$ we say that μ satisfies a "δ-upper threshold" with constant ϱ_1 if

$$\sup\{\mathbb{E}_{\mu^N}[\mu(K_N)] : N \leqslant \exp(\varrho_1 n)\} \leqslant \delta \tag{3.1}$$

and that μ satisfies a "δ-lower threshold" with constant ϱ_2 if

$$\inf\{\mathbb{E}_{\mu^N}[\mu(K_N)] : N \geqslant \exp(\varrho_2 n)\} \geqslant 1-\delta. \tag{3.2}$$

Then, we define the functions $\varrho_1(\mu,\delta) = \sup\{\varrho_1 :$ (3.1) holds true$\}$ and $\varrho_2(\mu,\delta) = \inf\{\varrho_2 :$ (3.2) holds true$\}$. Our main goal is to obtain upper bounds for the difference

$$\varrho(\mu,\delta) := \varrho_2(\mu,\delta) - \varrho_1(\mu,\delta)$$

for any fixed $\delta \in \left(0, \frac{1}{2}\right)$.

One may also consider a sequence $\{\mu_n\}_{n=1}^{\infty}$ of log-concave probability measures μ_n on $\mathbb{R}^n$. Then, we say that $\{\mu_n\}_{n=1}^{\infty}$ exhibits a "sharp threshold" if there exists a sequence $\{\delta_n\}_{n=1}^{\infty}$ of positive reals such that $\delta_n \to 0$ and $\varrho(\mu_n,\delta_n) \to 0$ as $n\to\infty$.

We present a general approach to the problem, working with an arbitrary log-concave probability measure μ on $\mathbb{R}^n$, which is based on the Cramer transform of μ. Recall that the logarithmic Laplace transform of μ is defined by

$$\Lambda_\mu(\xi) = \ln\Big(\int_{\mathbb{R}^n} e^{\langle \xi, z\rangle} d\mu(z)\Big), \qquad \xi\in\mathbb{R}^n$$

and the Cramer transform of μ is the Legendre transform of Λ_μ, defined by

$$\Lambda_\mu^*(x) = \sup_{\xi\in\mathbb{R}^n} \left\{\langle x,\xi\rangle - \Lambda_\mu(\xi)\right\}, \qquad x\in\mathbb{R}^n.$$

For every $t>0$ we set

$$B_t(\mu) := \{x\in\mathbb{R}^n : \Lambda_\mu^*(x)\leqslant t\}.$$

From the definition of Λ_μ^* one can easily check that for every $x\in\mathbb{R}^n$ we have $\varphi_\mu(x) \leqslant \exp(-\Lambda_\mu^*(x))$. In particular, for any $t>0$ and for all $x\notin B_t(\mu)$ we have that $\varphi_\mu(x)\leqslant \exp(-t)$. A main idea, which appears in all the previous works on this topic, is to show that φ_μ is almost constant on the boundary $\partial(B_t(\mu))$ of $B_t(\mu)$. Our first main result shows that this is true, in general, if $\mu=\mu_K$ is the uniform measure on a centered convex body of volume 1 in $\mathbb{R}^n$.

Theorem 3.1. *Let K be a centered convex body of volume 1 in $\mathbb{R}^n$. Then, for every $t>0$ we have that*

$$\inf\{\varphi_{\mu_K}(x) : x\in B_t(\mu_K)\}\geqslant \frac{1}{10}\exp(-t-2\sqrt{n}).$$

This implies that

$$\omega_{\mu_K}(x)-5\sqrt{n}\leqslant \Lambda^*(x)\leqslant \omega_{\mu_K}(x)$$

for every $x\in\mathbb{R}^n$, where $\omega_{\mu_K}(x)=\ln\left(\frac{1}{\varphi_{\mu_K}(x)}\right)$.

Theorem 3.1 may be viewed as a version of Cramér's theorem (see [9]) for random vectors uniformly distributed in convex bodies. Its proof exploits techniques from the theory of large deviations and a theorem of Nguyen [17] which is exactly the ingredient that forces us to consider only uniform measures on convex bodies. It seems harder to prove, if true, an analogous estimate for any centered log-concave probability measure μ on $\mathbb{R}^n$.

The second step in our approach is to consider, for any centered log-concave probability measure μ on $\mathbb{R}^n$, the parameter

$$\beta(\mu)=\frac{\mathrm{Var}_\mu(\Lambda_\mu^*)}{(\mathbb{E}_\mu(\Lambda_\mu^*))^2} \tag{3.3}$$

provided that

$$\|\Lambda_\mu^*\|_{L^2(\mu)}=\big(\mathbb{E}_\mu\big((\Lambda_\mu^*)^2\big)\big)^{1/2}<\infty.$$

Roughly speaking, the plan is the following: provided that φ_μ is "almost constant" on $\partial(B_t(\mu))$ for all $t>0$ and that $\beta(\mu)=o_n(1)$, one can establish a "sharp threshold" for the expected measure of K_N with

$$\varrho_2\approx\varrho_1\approx\|\Lambda_\mu^*\|_{L^1(\mu)}=\mathbb{E}_\mu(\Lambda_\mu^*).$$

Note that it is not clear in advance that Λ_μ^* has bounded second or higher order moments, which is necessary so that $\beta(\mu)$ would be well-defined. We obtain an affirmative answer in the case of the uniform measure on a convex body. In fact we cover the more general case of κ-concave probability measures, $\kappa\in(0,1/n]$, which are supported on a centered convex body.

Theorem 3.2. *Let K be a centered convex body of volume 1 in $\mathbb{R}^n$. Let $\kappa\in(0,1/n]$ and let μ be a centered κ-concave probability measure with $\mathrm{supp}(\mu)=K$. Then,*

$$\int_{\mathbb{R}^n} e^{\frac{\kappa\Lambda_\mu^*(x)}{2}}\,d\mu(x)<\infty.$$

In particular, for all $p\geqslant 1$ we have that $\mathbb{E}_\mu\big((\Lambda_\mu^*(x))^p\big)<\infty$.

The method of proof of Theorem 3.2 gives in fact reasonable upper bounds for $\|\Lambda_\mu^*\|_{L^p(\mu)}$. In particular, if we assume that $\mu=\mu_K$ is the uniform measure on a centered convex body then we obtain a sharp two sided estimate for the most interesting case where $p=1$ or 2.

Theorem 3.3. *Let K be a centered convex body of volume 1 in $\mathbb{R}^n$, $n \geqslant 2$. Then,*

$$c_1 n/L^2_{\mu_K} \leqslant \|\Lambda^*_{\mu_K}\|_{L^1(\mu_K)} \leqslant \|\Lambda^*_{\mu_K}\|_{L^2(\mu_K)} \leqslant c_2 n \ln n,$$

where L_{μ_K} is the isotropic constant of the uniform measure μ_K on K and $c_1, c_2 > 0$ are absolute constants.

The left-hand side inequality of Theorem 3.3 follows easily from Theorem 2.1, one of the main results in [4]. Both the lower and the upper bound are of optimal order with respect to the dimension. This can be seen e.g. from the example of the uniform measure on the cube or the Euclidean ball, respectively.

Besides Theorem 3.2, we show that Λ^*_μ has finite moments of all orders in the following cases:

(i) If μ is a centered probability measure on $\mathbb{R}$ which is absolutely continuous with respect to Lebesgue measure or a product of such measures.

(ii) If μ is a centered log-concave probability measure on $\mathbb{R}^n$ and there exists a function $g : [1, \infty) \to [1, \infty)$ with $\lim_{t\to\infty} g(t)/\ln(t+1) = +\infty$ such that $Z^+_t(\mu) \supseteq g(t) Z^+_2(\mu)$ for all $t \geqslant 2$, where $\{Z^+_t(\mu)\}_{t\geqslant 1}$ is the family of one-sided L_t-centroid bodies of μ.

Again, it seems harder to prove, if true, an analogous result for any centered log-concave probability measure μ on $\mathbb{R}^n$.

Next we show how one can use the previous results to obtain bounds for $\varrho(\mu, \delta)$. We also clarify the role of the parameter $\beta(\mu)$. One would hope that $\beta(\mu)$ is small as the dimension increases, e.g. $\beta(\mu) \leqslant c/\sqrt{n}$. If so, then the next general result provides satisfactory lower bounds for $\varrho_1(\mu, \delta)$.

Theorem 3.4. *Let μ be a centered log-concave probability measure on $\mathbb{R}^n$. Assume that $\beta(\mu) < 1/8$ and $8\beta(\mu) < \delta < 1$. If $n/L^2_\mu \geqslant c_2 \ln(2/\delta)\sqrt{\delta/\beta(\mu)}$ where L_μ is the isotropic constant of μ, then*

$$\varrho_1(\mu, \delta) \geqslant \left(1 - \sqrt{8\beta(\mu)/\delta}\right) \frac{\mathbb{E}_\mu(\Lambda^*_\mu)}{n}.$$

We are able to give satisfactory upper bounds for $\varrho_2(\mu, \delta)$ in the case where $\mu = \mu_K$ is the uniform measure on a centered convex body K of volume 1 in $\mathbb{R}^n$.

Theorem 3.5. *Let K be a centered convex body of volume 1 in $\mathbb{R}^n$. Let $\beta(\mu_K) < 1/2$ and $2\beta(\mu_K) < \delta < 1$. If $n/L^2_{\mu_K} \geqslant c_2 \ln(2/\delta)\sqrt{\delta/\beta(\mu_K)}$ then*

$$\varrho_2(\mu_K, \delta) \leqslant \left(1 + \sqrt{8\beta(\mu_K)/\delta}\right) \frac{\mathbb{E}_{\mu_K}(\Lambda^*_{\mu_K})}{n}.$$

Combining these two results we see that, provided that $\beta(\mu_K)$ is small compared to a fixed $\delta \in (0,1)$, we have a threshold of the order

$$\varrho(\mu_K, \delta) \leqslant \frac{c}{n}\sqrt{\frac{\mathrm{Var}_{\mu_K}(\Lambda^*_{\mu_K})}{\delta}}.$$

The above discussion leaves open the question to estimate

$$\beta_n := \sup\{\beta(\mu) : \mu \text{ is a centered log-concave probability measure on } \mathbb{R}^n\}.$$

We illustrate the method that we develop in this work with a number of examples. We consider first the standard examples of the uniform measure on the unit cube and the Gaussian measure. As a direct consequence of our results, in both cases we obtain a bound $\varrho(\mu,\delta)\leqslant c(\delta)/\sqrt{n}$ for the threshold, where $c(\delta)>0$ is a constant depending on δ. Finally, we examine the case of the uniform measure on the Euclidean ball D_n of volume 1 in $\mathbb{R}^n$. Here, we exploit the general fact that, for any centered convex body K of volume 1 in $\mathbb{R}^n$, if we consider the function $\omega_{\mu_K}(x)=\ln(1/\varphi_{\mu_K}(x))$ and the parameter

$$\tau(\mu_K)=\frac{\mathrm{Var}_\mu(\omega_{\mu_K})}{(\mathbb{E}_\mu(\omega_{\mu_K}))^2}$$

then

$$\beta(\mu_K)=\left(\tau(\mu_K)+O(L_{\mu_K}^2/\sqrt{n})\right)\left(1+O(L_{\mu_K}^2/\sqrt{n})\right).$$

In the case of the ball D_n, working with $\omega_{\mu_{D_n}}$ is easier than working with $\Lambda^*_{\mu_{D_n}}$, and we can check that

$$\tau(\mu_{D_n})=O(1/n).$$

This leads to the following sharp threshold for the Euclidean ball.

Theorem 3.6. *Let D_n be the centered Euclidean ball of volume 1 in $\mathbb{R}^n$. Then, the sequence $\mu_n:=\mu_{D_n}$ exhibits a sharp threshold with $\varrho(\mu_n,\delta)\leqslant\frac{c}{\sqrt{\delta n}}$ and e.g. in the case where n is even we have that*

$$\mathbb{E}_{\mu_n}(\Lambda^*_{\mu_n})=\frac{(n+1)}{2}H_{\frac{n}{2}}+O(\sqrt{n})$$

as $n\to\infty$, where $H_m=\sum_{k=1}^m\frac{1}{k}$.

References

[1] Mathoverflow, *Expected value of Tukey's half-space depth for log-concave measures*, Question asked in `https://mathoverflow.net/questions/386972/`.

[2] G. Bonnet, G. Chasapis, J. Grote, D. Temesvari and N. Turchi, Threshold phenomena for high-dimensional random polytopes, Commun. Contemp. Math. **21** (2019), no. 5, 1850038, 30 pp.

[3] G. Bonnet, Z. Kabluchko and N. Turchi, *Phase transition for the volume of high-dimensional random polytopes*, Random Structures Algorithms **58** (2021), no. 4, 648–663.

[4] S. Brazitikos, A. Giannopoulos and M. Pafis, *Half-space depth of log-concave probability measures*, Preprint.

[5] S. Brazitikos, A. Giannopoulos and M. Pafis, *Threshold for the expected measure of random polytopes*, Preprint.

[6] S. Brazitikos, A. Giannopoulos, P. Valettas and B-H. Vritsiou, *Geometry of isotropic convex bodies*, Mathematical Surveys and Monographs **196**, American Mathematical Society, Providence, RI, 2014.

[7] D. Chakraborti, T. Tkocz and B-H. Vritsiou, *A note on volume thresholds for random polytopes*, Geom. Dedicata **213** (2021), 423–431.

[8] Y. Chen, *An almost constant lower bound of the isoperimetric coefficient in the KLS conjecture*, Geom. Funct. Anal. **31** (2021), 34-61.

[9] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, Springer, New York, 1998.

[10] D. L. Donoho and M. Gasko, *Breakdown properties of location estimates based on halfspace depth and projected outlyingness*, Ann. Statist. **20** (1982), 1803–1827.

[11] M. E. Dyer, Z. Furedi and C. McDiarmid, *Volumes spanned by random points in the hypercube*, Random Structures Algorithms **3** (1992), 91–106.

[12] A. Frieze, W. Pegden and T. Tkocz, *Random volumes in d-dimensional polytopes*, Discrete Anal. 2020, Paper No. 15, 17 pp.

[13] D. Gatzouras and A. Giannopoulos, *Threshold for the volume spanned by random points with independent coordinates*, Israel J. Math. **169** (2009), 125–153.

[14] J. L. Hodges Jr., *A bivariate sign test*, Ann. Math. Statist. **26** (1955), 523–527.

[15] B. Klartag and J. Lehec, *Bourgain's slicing problem and KLS isoperimetry up to polylog*, Preprint.

[16] S. Nagy, C. Schutt and E. M. Werner, *Halfspace depth and floating body*, Stat. Surv. **13** (2019), 52–118.

[17] V. H. Nguyen, *Dimensional variance inequalities of Brascamp-Lieb type and a local approach to dimensional Prékopa's theorem*, J. Funct. Anal. **266** (2014), 931–955.

[18] P. Pivovarov, *Volume thresholds for Gaussian and spherical random polytopes and their duals*, Studia Math. **183** (2007), no. 1, 15–34.

[19] C. G. Small, *Measures of centrality for multivariate and directional distributions*, Canad. J. Statist. **15** (1987), 31–39.

[20] J. W. Tukey, *Mathematics and the picturing of data*, In Proceedings of the International Congress of Mathematicians (Vancouver, B. C., 1974), Vol. 2 523–531. Canad. Math. Congress, Montreal, Que.

Silouanos Brazitikos: Department of Mathematics, National and Kapodistrian University of Athens, Panepistimioupolis 157-84, Athens, Greece.

E-mail: silouanb@math.uoa.gr

Apostolos Giannopoulos: Department of Mathematics, National and Kapodistrian University of Athens, Panepistimioupolis 157-84, Athens, Greece.

E-mail: apgiannop@math.uoa.gr

Minas Pafis: Department of Mathematics, National and Kapodistrian University of Athens, Panepistimioupolis 157-84, Athens, Greece.

E-mail: mipafis@math.uoa.gr

Μαθαίνοντας πολυώνυμα στον διακριτό κύβο

*Αλέξανδρος Εσκενάζης**

Οι ανισότητες των Littlewood και Bohnenblust–Hille

Έστω $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ και $a_1, a_2, \ldots \in \mathbb{K}$. Τετριμμένα, έχουμε ότι

$$\sum_{i\geq 1} |a_i| = \max\Big\{\Big|\sum_{i\geq 1} a_i x_i\Big| : \max_{i\geq 1} |x_i| \leq 1\Big\}.$$

Με αφορμή την παραπάνω ταυτότητα, ο Daniell έθεσε την εξής ερώτηση.

Ερώτηση. Υπάρχει σταθερά $C < \infty$ ώστε για κάθε $\{a_{ij}\}_{i,j\geq 1} \subset \mathbb{K}$ να έχουμε

$$\sum_{i,j\geq 1} |a_{ij}| \leq C \sup\Big\{\Big|\sum_{i,j\geq 1} a_{ij} x_i y_j\Big| : \|x\|_{\ell_\infty}, \|y\|_{\ell_\infty} \leq 1\Big\} \;;$$

Πρέπει να σημειωθεί ότι δεν τίθεται περιορισμός στο πλήθος των a_{ij}, οπότε αναζητούμε μια ανισότητα στην οποία η σταθερά είναι *ανεξάρτητη της διάστασης*. Στη διάσημη εργασία [11] ο Littlewood απάντησε αρνητικά στην ερώτηση του Daniell και απέδειξε το εξής σημαντικό αποτέλεσμα.

Η $\frac{4}{3}$-ανισότητα του Littlewood. Υπάρχει σταθερά $C < \infty$ ώστε για κάθε $\{a_{ij}\}_{i,j\geq 1}$,

$$\Big(\sum_{i,j\geq 1} |a_{ij}|^{4/3}\Big)^{3/4} \leq C \sup\Big\{\Big|\sum_{i,j\geq 1} a_{ij} x_i y_j\Big| : \|x\|_{\ell_\infty}, \|y\|_{\ell_\infty} \leq 1\Big\}.$$

Επιπλέον, ο εκθέτης $\frac{4}{3}$ είναι ο ελάχιστος για τον οποίο η παραπάνω ανισότητα ισχύει.

Η ανισότητα του Littlewood είναι ειδική περίπτωση μιας πολύ γενικότερης ανισότητας για πολυώνυμα που μελετήθηκε για πρώτη φορά από τους Bohnenblust και Hille [2]. Αν $\alpha = (\alpha_1, \alpha_2, \ldots)$ είναι μια τελικά μηδενική ακολουθία φυσικών αριθμών και $x = (x_1, x_2, \ldots)$, τότε θα χρησιμοποιούμε τους συμβολισμούς $|\alpha| = |\alpha_1| + |\alpha_2| + \cdots$ και $x^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \cdots$.

Πολυωνυμική ανισότητα Bohnenblust–Hille. Για κάθε $d \in \mathbb{N}$, υπάρχει σταθερά $B_d^{\mathbb{K}} < \infty$ ώστε για κάθε πολυώνυμο της μορφής

$$\forall\, x = (x_1, x_2, \ldots), \qquad p(x) = \sum_{|\alpha|\leq d} c_\alpha x^\alpha$$

*Το παρόν είναι η περίληψη μιας ομιλίας που δόθηκε στο συνέδριο στη μνήμη του Δ. Γατζούρα (Πανεπιστήμιο Αθηνών, 12-13 Μαρτίου 2022) και βασίζεται στις κοινές εργασίες [6, 7] με τους P. Ivanisvili και L. Streck.

να ισχύει

$$\Big(\sum_{|\alpha|\le d}|c_\alpha|^{\frac{2d}{d+1}}\Big)^{\frac{d+1}{2d}}\le B_d^{\mathbb{K}}\sup\big\{|p(x)|:\ \|x\|_{\ell_\infty}\le 1\big\}.$$

Επιπλέον, ο εκθέτης $\frac{2d}{d+1}$ είναι βέλτιστος.

Η ακριβής ασυμπτωτική συμπεριφορά των σταθερών $B_d^{\mathbb{R}}$ και $B_d^{\mathbb{C}}$ παραμένει άγνωστη. Παρ'όλα αυτά γνωρίζουμε ότι οι δύο είναι πολύ διαφορετικές μεταξύ τους, ειδικότερα ότι $B_d^{\mathbb{C}}\le C^{\sqrt{d\log d}}$ για μια σταθερά $C>1$ ενώ $\limsup_{d\to\infty}(B_d^{\mathbb{R}})^{1/d}=1+\sqrt{2}$ (ενδεικτικά αναφέρουμε τα [4, 1, 3, 5]).

Η βάση Walsh

Για έναν φυσικό αριθμό $n\in\mathbb{N}$ και ένα υποσύνολο $S\subseteq\{1,\ldots,n\}$ η αντίστοιχη συνάρτηση Walsh $w_S:\{-1,1\}^n\to\{-1,1\}$ ορίζεται ως $w_S(x)=\prod_{i\in S}x_i$, όπου $x=(x_1,\ldots,x_n)\in\{-1,1\}^n$. Οι 2^n το πλήθος συναρτήσεις $\{w_S\}_{S\subseteq\{1,\ldots,n\}}$ είναι ανά δύο κάθετες στο χώρο L_2 που ορίζεται από το ομοιόμορφο μέτρο πιθανότητας στο $\{-1,1\}^n$ και συνεπώς κάθε συνάρτηση $f:\{-1,1\}^n\to\mathbb{R}$ μπορεί να αναπτυχθεί μοναδικά ως

$$\forall\, x\in\{-1,1\}^n,\qquad f(x)=\sum_{S\subseteq\{1,\ldots,n\}}\hat{f}(S)w_S(x),$$

όπου

$$\hat{f}(S)=\mathbb{E}_x[f(x)w_S(x)].$$

Λέμε ότι η συνάρτηση f έχει βαθμό το πολύ d και γράφουμε $\deg(f)\le d$ αν $\hat{f}(S)=0$ για κάθε σύνολο $S\subseteq\{1,\ldots,n\}$ με $|S|>d$. Το ανάπτυγμα Walsh είναι εξαιρετικά χρήσιμο στην ανάλυση στο διακριτό κύβο. Σε αυτή την ομιλία θα δούμε μερικές εφαρμογές του στην υπολογιστική θεωρία μάθησης (computational learning theory).

Μαθαίνοντας κλάσεις συναρτήσεων στο διακριτό κύβο

Έστω $n\in\mathbb{N}$, $\mathcal{F}$ μια κλάση πραγματικών συναρτήσεων στο $\{-1,1\}^n$ και $f\in\mathcal{F}$ μια άγνωστη συνάρτηση. Θα μας απασχολήσει το εξής υπολογιστικό πρόβλημα. Ας υποθέσουμε ότι $X_1,X_2,\ldots$ είναι μια ακολουθία ανεξάρτητων τυχαίων μεταβλητών, καθεμία εκ των οποίων είναι ομοιόμορφα κατανεμημένη στο $\{-1,1\}^n$.

Αναζητούμε το ελάχιστο $N\in\mathbb{N}$ ώστε γνωρίζοντας τα δεδομένα

$$(X_1,f(X_1)),\ldots,(X_N,f(X_N)),$$

να μπορούμε να κατασκευάσουμε μια *τυχαία* $h:\{-1,1\}^n\to\mathbb{R}$ η οποία να είναι μια καλή προσέγγιση της f με μεγάλη πιθανότητα. Ειδικότερα, σε αυτή την περίληψη θα ζητήσουμε η συνάρτηση h να ικανοποιεί $\|h-f\|_{L_2}^2\le\varepsilon$ με πιθανότητα τουλάχιστον $1-\delta$ όπου οι $\varepsilon,\delta\in(0,1)$ είναι δυο προεπιλεγμένες παράμετροι.

Ο Αλγόριθμος Χαμηλού Βαθμού

Το παραπάνω πρόβλημα έχει μελετηθεί στη βιβλιογραφία για πολλές διαφορετικές κλάσεις $\mathcal{F}$ συναρτήσεων. Το πρώτο σχετικό αυστηρό μαθηματικό αποτέλεσμα είναι ο διάσημος αλγόριθμος (Low-Degree Algorithm) των Linial, Mansour και Nisan [10] οι οποίοι μελέτησαν την κλάση

$$\mathcal{F}_{n,d} = \big\{ f : \{-1,1\}^n \to [-1,1] :\ \deg(f) \leq d \big\}.$$

Ο αλγόριθμός τους δίνει το εξής πολυωνυμικό φράγμα για την παράμετρο N.

Ο Αλγόριθμος Χαμηλού Βαθμού. Έστω $n \in \mathbb{N}$, $d \in \{1,\ldots,n\}$ και $\varepsilon, \delta \in (0,1)$. Υπάρχει ένας αλγόριθμος που μαθαίνει κάθε $f \in \mathcal{F}_{n,d}$ με πιθανότητα τουλάχιστον $1-\delta$ και L_2-σφάλμα το πολύ ε από το πολύ N δείγματα, όπου $N \leq \frac{2n^d}{\varepsilon} \log\big(\frac{2n^d}{\delta}\big)$.

Απόδειξη. Η ιδέα είναι να πάρουμε αρκετά δείγματα ώστε να μπορέσουμε να υπολογίσουμε με μεγάλη ακρίβεια όλους τους συντελεστές στο ανάπτυγμα Walsh της f. Για ένα υποσύνολο $S \subseteq \{1,\ldots,n\}$ με $|S| \leq d$ θεωρούμε τον εμπειρικό συντελεστη Walsh

$$\alpha_S = \frac{1}{N} \sum_{j=1}^{N} f(X_j) w_S(X_j).$$

Η τυχαία μεταβλητή α_S είναι ένα άθροισμα ανεξάρτητων φραγμένων τυχαίων μεταβλητών και ικανοποιεί $\mathbb{E}[\alpha_S] = \hat{f}(S)$, οπότε από την ανισότητα Hoeffding έχουμε την εκτίμηση

$$\forall\, b > 0, \qquad \mathbb{P}\big\{ |\alpha_S - \hat{f}(S)| \geq b \big\} \leq 2e^{-Nb^2/2}.$$

Συνεπώς, αν ορίσουμε

$$N = N(b) = \left\lceil \frac{2}{b^2} \log\left(\frac{2}{\delta} \sum_{k=0}^{d} \binom{n}{k} \right) \right\rceil,$$

θα έχουμε

$$\mathbb{P}\big\{ |\alpha_S - \hat{f}(S)| < b, \forall\, S \big\} \geq 1 - 2\sum_{k=0}^{d} \binom{n}{k} e^{-Nb^2/2} \geq 1 - \delta$$

από την επιλογή του N. Έστω G_b το ενδεχόμενο στο αριστερό μέλος αυτής της ανισότητας και θεωρούμε την τυχαία συνάρτηση $h : \{-1,1\}^n \to \mathbb{R}$ που ορίζεται ως

$$\forall\, x \in \{-1,1\}^n, \qquad h(x) = \sum_{S \subseteq \{1,\ldots,n\}} \alpha_S w_S(x).$$

Τότε, από την ταυτότητα Parseval, στο ενδεχόμενο G_b έχουμε

$$\|h - f\|_{L_2}^2 = \sum_S (\alpha_S - \hat{f}(S))^2 < \sum_{k=0}^{d} \binom{n}{k} b^2 \leq \varepsilon$$

για $b^2 \leq \varepsilon / \sum_{k=0}^{d} \binom{n}{k}$. Αντικαθιστώντας αυτή την τιμή του b στο N έχουμε το αποτέλεσμα. □

Μια καλύτερη εκτίμηση

Η εκτίμηση των Linial, Mansour και Nisan ήταν το μόνο γνωστό φράγμα για αυτή την ποσότητα μέχρι που πρόσφατα οι Iyer, Rao, Reis, Rothvoss και Yehudayoff [8] απέδειξαν πως $O_{d,\varepsilon,\delta}(n^{d-1}\log n)$ δείγματα αρκούν βασιζόμενοι σε μια ιδέα των Kushilevitz και Mansour [9]. Το κύριο αποτέλεσμα αυτής της ομιλίας είναι μια λογαριθμική εκτίμηση από το [6]. Θα συμβολίζουμε με $B_d^{\{\pm1\}}$ την ελάχιστη σταθερά για την οποία ισχύει το εξής: για $n \geq d$ και κάθε συνάρτηση $f : \{-1,1\}^n \to \mathbb{R}$ βαθμού το πολύ d, έχουμε

$$\Big(\sum_{|S|\leq d} |\hat{f}(S)|^{\frac{2d}{d+1}}\Big)^{\frac{d+1}{2d}} \leq B_d^{\{\pm1\}}\|f\|_{L_\infty}.$$

Ταυτίζοντας την f με την πλειογραμμική της επέκταση στο συνεχή κύβο $[-1,1]^n$, έπεται ότι

$$\max_{x\in[-1,1]^n} |f(x)| = \max_{x\in\{-1,1\}^n} |f(x)|$$

και συνεπώς $B_d^{\{\pm1\}} \leq B_d^{\mathbb{R}}$. Είναι γνωστό [5] ότι υπάρχει $C > 1$ ώστε $B_d^{\{\pm1\}} \leq C^{\sqrt{d\log d}}$.

Θεώρημα. Έστω $\varepsilon, \delta \in (0,1)$, $n \in \mathbb{N}$, $d \in \{1,\ldots,n\}$ και μια συνάρτηση $f \in \mathcal{F}_{n,d}$. Αν το $N \in \mathbb{N}$ ικανοποιεί

$$N \geq \frac{e^8 d^2}{\varepsilon^{d+1}}(B_d^{\{\pm1\}})^{2d}\log\Big(\frac{n}{\delta}\Big),$$

τότε N ανεξάρτητα και ομοιόμορφα τυχαία δείγματα της μορφής $(x, f(x))$, όπου $x \in \{-1,1\}^n$, αρκούν για να κατασκευαστεί μια τυχαία συνάρτηση $h : \{-1,1\}^n \to \mathbb{R}$ που ικανοποιεί $\|h-f\|_{L_2}^2 < \varepsilon$ με πιθανότητα τουλάχιστον $1-\delta$.

Απόδειξη. Όπως πριν, σταθεροποιούμε $b > 0$ και ορίζουμε

$$N = N(b) = \left\lceil \frac{2}{b^2}\log\left(\frac{2}{\delta}\sum_{k=0}^{d}\binom{n}{k}\right)\right\rceil,$$

ώστε αν

$$\alpha_S = \frac{1}{N}\sum_{j=1}^{N} f(X_j)w_S(X_j),$$

τότε το ενδεχόμενο G_b έχει πιθανότητα

$$\mathbb{P}(G_b) = \mathbb{P}\big\{|\alpha_S - \hat{f}(S)| < b, \forall\, S\big\} \geq 1 - 2\sum_{k=0}^{d}\binom{n}{k}e^{-Nb^2/2} \geq 1-\delta.$$

Έστω μια νέα παράμετρος $a > b$ και θεωρούμε την τυχαία οικογένεια συνόλων Σ_a που ορίζεται ως

$$\Sigma_a = \{S : \ |\alpha_S| > a\}$$

και την τυχαία συνάρτηση $h_{a,b} : \{-1,1\}^n \to \mathbb{R}$, όπου

$$\forall\, x \in \{-1,1\}^n, \qquad h_{a,b}(x) = \sum_{S\in\Sigma_a}\alpha_S w_S(x).$$

Στο ενδεχόμενο G_b έχουμε ότι αν $S \in \Sigma_a$, τότε $|\hat{f}(S)| > a - b$, ενώ αν $S \notin \Sigma_a$, τότε $|\hat{f}(S)| < a + b$. Συνεπώς, συμπεραίνουμε από την ανισότητα Bohnenblust–Hille στο διακριτό κύβο ότι

$$|\Sigma_a| \leq (a-b)^{-\frac{2d}{d+1}} \sum_{S \in \Sigma_a} |\hat{f}(S)|^{\frac{2d}{d+1}} \leq (B_d^{\{\pm 1\}})^{\frac{2d}{d+1}} (a-b)^{-\frac{2d}{d+1}}.$$

Έτσι, συνδυάζοντας την ταυτότητα του Parseval με μια ακόμη εφαρμογή της ανισότητας Bohnenblust–Hille καταλήγουμε πως η τυχαία συνάρτηση $h_{a,b}$ ικανοποιεί

$$\begin{aligned}\|h_{a,b} - f\|_{L_2}^2 &= \sum_{S \in \Sigma_a} (\alpha_S - \hat{f}(S))^2 + \sum_{S \notin \Sigma_a} \hat{f}(S)^2 < b^2 |\Sigma_a| + (a+b)^{\frac{2}{d+1}} \sum_{S \notin \Sigma_a} |\hat{f}(S)|^{\frac{2d}{d+1}} \\ &\leq (B_d^{\{\pm 1\}})^{\frac{2d}{d+1}} \big(b^2 (a-b)^{-\frac{2d}{d+1}} + (a+b)^{\frac{2}{d+1}} \big).\end{aligned}$$

Διαλέγοντας τις βέλτιστες παραμέτρους a, b για τις οποίες το δεξί μέλος ειναι μικρότερο του ε, έπεται το συμπέρασμα του θεωρήματος. $\square$

Κάτω φράγματα

Η πρόσφατη εργασία [7] περιλαμβάνει κάτω φράγματα για τον ελάχιστο αριθμό δειγμάτων που απαιτούνται για να μάθει κανείς μια γενική συνάρτηση $f \in \mathcal{F}_{n,d}$. Το κεντρικό αποτέλεσμα του άρθρου δείχνει πως $\Omega_{d,\varepsilon,\delta}(\log n)$ δείγματα είναι απαραίτητα για αυτό το σκοπό, κάτι που αποδεικνύει πως το θεώρημα της προηγούμενης ενότητας είναι βέλτιστο καθώς $n \to \infty$. Η ιδέα είναι πως αν μια κλάση $\mathcal{F}$ περιλαμβάνει k Boolean συναρτήσεις που απέχουν πολύ ανά δύο, τότε κανένας αλγόριθμος δεν μπορεί να τις ξεχωρίσει χρησιμοποιώντας λιγότερα από $\Omega(\log k)$ δείγματα. Συνεπώς, το αποτέλεσμα έπεται επειδή η *μετρική εντροπία* της κλάσης $(\mathcal{F}_{n,d}, \|\cdot\|_{L_2})$ συμπεριφέρεται λογαριθμικά στο n καθώς $n \to \infty$. Το κύριο τεχνικό αποτέλεσμα του [7] είναι μια νέα εκτίμηση της μετρικής εντροπίας αυτής της κλάσης που είναι βέλτιστη καθώς $n \to \infty$ *και* καθώς $d \to \infty$, για $d \leq \log_2 n$.

References

[1] F. Bayart, D. Pellegrino, and J. B. Seoane-Sepúlveda, *The Bohr radius of the n-dimensional polydisk is equivalent to $\sqrt{(\log n)/n}$*, Adv. Math., 264 (2014), pp. 726-746.

[2] H. F. Bohnenblust and E. Hille, *On the absolute convergence of Dirichlet series*, Ann. of Math. (2), 32 (1931), pp. 600-622.

[3] J. R. Campos, P. Jiménez-Rodriguez, G. A. Muñoz-Fernández, D. Pellegrino, and J. B. Seoane-Sepúlveda, *On the real polynomial Bohnenblust-Hille inequality*, Linear Algebra Appl., 465 (2015), pp. 391-400.

[4] A. Defant, L. Frerick, J. Ortega-Cerdà, M. Ounaies, and K. Seip, *The Bohnenblust-Hille inequality for homogeneous polynomials is hypercontractive*, Ann. of Math. (2), 174 (2011), pp. 485-497.

[5] A. Defant, M. Mastyło, and A. Pérez, *On the Fourier spectrum of functions on Boolean cubes*, Math. Ann., 374 (2019), pp. 653-680.

[6] A. Eskenazis and P. Ivanisvili, *Learning low-degree functions from a logarithmic number of random queries*. To appear in Proceedings of STOC 2022, 2021.

[7] A. Eskenazis, P. Ivanisvili, and L. Streck, *Low-degree learning and the metric entropy of polynomials*. Preprint, 2022.

[8] S. Iyer, A. Rao, V. Reis, T. Rothvoss, and A. Yehudayoff, *Tight bounds on the Fourier growth of bounded functions on the hypercube*. To appear in ECCC 2021, 2021.

[9] E. Kushilevitz and Y. Mansour, *Learning decision trees using the Fourier spectrum*, SIAM J. Comput., 22 (1993), pp. 1331-1348.

[10] N. Linial, Y. Mansour, and N. Nisan, *Constant depth circuits, Fourier transform, and learnability*, J. Assoc. Comput. Mach., 40 (1993), pp. 607-620.

[11] J. E. Littlewood, *On bounded bilinear forms in an infinite number of variables*, Q. J. Math., os-1 (1930), pp. 164-174.

Alexandros Eskenazis: CNRS, Sorbonne Université, France and Trinity College, University of Cambridge, UK.

E-mail: alexandros.eskenazis@imj-prg.fr; ae466@cam.ac.uk

Joint ergodicity of sequences - An exposition

*Nikos Frantzikinakis**

Abstract

A collection of integer sequences is jointly ergodic if for every ergodic measure preserving system the multiple ergodic averages, with iterates given by this collection of sequences, converge in the mean to the product of the integrals of the functions involved. Convenient necessary and sufficient conditions for joint ergodicity were given in [11] and this exposition uses a simplified version of the argument in [11] in order to recover its main results under somewhat stronger assumptions. The argument we give is rather short and avoids deep tools from ergodic theory. The main result can be used to prove new ergodic theorems and give vast simplifications of older results that depended on deep machinery from ergodic theory.

1 Introduction

The polynomial Szemerédi theorem of Bergelson and Leibman [1] states that if Λ is a set of integers with positive upper density and $p_1, \ldots, p_\ell \in \mathbb{Z}[t]$ are polynomials with zero constant term, then there exist $m, n \in \mathbb{N}$ such that

$$m, m + p_1(n), \ldots, m + p_\ell(n) \in \Lambda.$$

This generalizes the theorem of Szemerédi [28] on arithmetic progressions that corresponds to the case where $p_1(n) = n, p_2(n) = 2n, \ldots, p_\ell(n) = \ell n$. The proof of Bergelson and Leibman uses ergodic theory and up to this day it is the only proof that covers the full generality of this result. Using the correspondence principle of Furstenberg [14, 15] it turns out that it suffices to verify the following: For every measure preserving system $(X, \mathcal{X}, \mu, T)$ and set $A \in \mathcal{X}$ with positive measure, there exists $n \in \mathbb{N}$ such that

$$\mu(A \cap T^{-p_1(n)}A \cap \cdots \cap T^{-p_\ell(n)}A) > 0.$$

The proof of this multiple recurrence property proceeds by analyzing the limiting behavior in $L^2(\mu)$ of the following multiple ergodic averages (see our notation for averages in Section 1.1)

$$\mathbb{E}_{n\in[N]}\, T^{p_1(n)}f_1 \cdot \ldots \cdot T^{p_\ell(n)}f_\ell. \tag{1.1}$$

Finding an explicit formula for this limit for all polynomials is still an unresolved problem, but in some cases the limit takes a particularly simple form, namely, it is the product of the integrals of the functions $f_1, \ldots, f_\ell$. Due to congruence obstructions this can only be the case for totally ergodic systems, which is the reason why we are particularly interested in this class of systems. The prototypical result was established

*The author was supported by the Hellenic Foundation for Research and Innovation, Project No: 1684.

by Furstenberg and Weiss [16] and states that in a totally ergodic system for every $f, g \in L^\infty(\mu)$ we have

$$\lim_{N\to\infty} \mathbb{E}_{n\in[N]} T^n f \cdot T^{n^2} g = \int f\, d\mu \cdot \int g\, d\mu \tag{1.2}$$

in $L^2(\mu)$. The proof of this result is rather involved and its most difficult component is the analysis of a special class of two step distal systems, called Conze-Lesigne systems (introduced in [7]), that control the limiting behavior of these averages. Conze-Lesigne systems are particular examples of systems with nilpotent structure, a concept that has played an important role in subsequent developments in the field. By combining the Host-Kra theory of characteristic factors [17] and equidistribution results on nilmanifolds from [22], the author and Kra extended in [13] the result of Furstenberg and Weiss by showing that in a totally ergodic system if the polynomials $p_1, \ldots, p_\ell \in \mathbb{Z}[t]$ are rationally independent,[1] then for all $f_1, \ldots, f_\ell \in L^\infty(\mu)$ we have

$$\lim_{N\to\infty} \mathbb{E}_{n\in[N]} T^{p_1(n)} f_1 \cdot \ldots \cdot T^{p_\ell(n)} f_\ell = \int f_1\, d\mu \cdots \int f_\ell\, d\mu \tag{1.3}$$

in $L^2(\mu)$. If the polynomials are rationally dependent, then easy examples of totally ergodic circle rotations show that the previous limit formula fails. In fact, when $p_1(n) = n, p_2(n) = 2n, \ldots, p_\ell(n) = \ell n$, the limit can be computed using the results in [17, 30, 31] and it turns out that it genuinely depends on the $(\ell-1)$-step nilsystems that are factors of the original system. So in order to obtain an explicit limit formula for the averages (1.1) for dependent polynomials, the use of deep structural results from ergodic theory and equidistribution results on nilmanifolds seems unavoidable. This is not the case though for rationally independent polynomials, and it has been a tantalizing open problem for quite a while to get an "elementary" proof for the limit formulas (1.2) and (1.3). The main purpose of this note is to reproduce a simplified version of an argument from [11] that accomplishes this goal. Moreover, as in [11], our main result (Theorem 2.1) gives a rather general statement that applies to a variety of sequences, not just polynomials, and this allows to prove some new convergence results and establish some conjectures. We record a few examples from recent literature in Section 2.3. As in [11], our argument was motivated by techniques of Peluse [24] and Peluse and Prendiville [26] of finitary nature that were originally devised to give quantitative estimates for special cases of the polynomial Szemerédi theorem.

1.1 Definitions and notation

With $\mathbb{N}$ we denote the set of positive integers and with $\mathbb{Z}_+$ the set of non-negative integers. For $t \in \mathbb{R}$ we let $e(t) := e^{2\pi i t}$. With $\mathbb{T}$ we denote the one dimensional torus and we often identify it with $\mathbb{R}/\mathbb{Z}$ or with $[0,1)$. With $\Re(z)$ we denote the real part of the complex number z. For $N \in \mathbb{N}$ we let $[N] := \{1, \ldots, N\}$. If $a\colon \mathbb{N}^s \to \mathbb{C}$ is a bounded sequence for some $s \in \mathbb{N}$ and A is a non-empty finite subset of $\mathbb{N}^s$, we let $\mathbb{E}_{n\in A}\, a(n) := \frac{1}{|A|}\sum_{n\in A} a(n)$.

[1] A collection of integer polynomials is *rationally independent*, if every non-trivial linear combination of the polynomials is non-constant.

2 Main results

2.1 Definitions

In order to facilitate our exposition we reproduce some definitions from [11].

Definition. We say that the collection of sequences $a_1, \ldots, a_\ell \colon \mathbb{N} \to \mathbb{Z}$ is *jointly ergodic for the ergodic system* $(X, \mathcal{X}, \mu, T)$, if for all $f_1, \ldots, f_\ell \in L^\infty(\mu)$ we have

$$\lim_{N\to\infty} \mathbb{E}_{n\in[N]} T^{a_1(n)} f_1 \cdot \ldots \cdot T^{a_\ell(n)} f_\ell = \int f_1 \, d\mu \cdot \ldots \cdot \int f_\ell \, d\mu \tag{2.1}$$

in $L^2(\mu)$.

If a collection of sequences is jointly ergodic for every ergodic system, then an ergodic decomposition argument shows that the limit formula (2.1) holds for every system $(X, \mathcal{X}, \mu, T)$ (not necessarily ergodic), if we use in place of the integrals $\int f_i \, d\mu$ the conditional expectations $\mathbb{E}(f_i|\mathcal{I}(T))$ ($\mathcal{I}(T)$ is the σ-algebra of T-invariant sets). This implies that the following strong multiple recurrence property holds

$$\lim_{N\to\infty} \mathbb{E}_{n\in[N]} \mu(A \cap T^{-a_1(n)} A \cap \cdots \cap T^{-a_\ell(n)} A) \geq (\mu(A))^{\ell+1},$$

for every system $(X, \mathcal{X}, \mu, T)$ and every $A \in \mathcal{X}$. It is then a consequence of the correspondence principle of Furstenberg [15] that every set of integers with positive upper density contains patterns of the form $m, m + a_1(n), \ldots, m + a_\ell(n)$, for some $m, n \in \mathbb{N}$.

Definition. If $(X, \mathcal{X}, \mu, T)$ is a system we defined its *spectrum* as follows

$$\mathrm{Spec}(T) := \{t \in [0, 1) \colon Tf = e(t)\, f \text{ for some non-zero } f \in L^2(\mu)\}.$$

For the definition of the seminorms $\|\!|\cdot|\!\|_s$ we refer the reader to Section 3.2.

Definition. We say that the collection of sequences $a_1, \ldots, a_\ell \colon \mathbb{N} \to \mathbb{Z}$ is:

(i) *good for seminorm estimates for the system* $(X, \mathcal{X}, \mu, T)$, if there exists $s \in \mathbb{N}$ such that if $f_1, \ldots, f_\ell \in L^\infty(\mu)$ and $\|\!|f_i|\!\|_s = 0$ for some $i \in \{1, \ldots, \ell\}$, then

$$\lim_{N\to\infty} \mathbb{E}_{n\in[N]} T^{a_1(n)} f_1 \cdot \ldots \cdot T^{a_\ell(n)} f_\ell = 0$$

in $L^2(\mu)$;

(ii) *good for equidistribution on* $S \subset [0, 1)$, if for all $t_1, \ldots, t_\ell \in S$, not all of them 0, we have

$$\lim_{N\to\infty} \mathbb{E}_{n\in[N]} e(a_1(n)t_1 + \cdots + a_\ell(n)t_\ell) = 0. \tag{2.2}$$

It is known [18, 23] that if $p_1, \ldots, p_\ell \colon \mathbb{N} \to \mathbb{Z}$ are polynomials with pairwise non-constant differences, then they are good for seminorm estimates for every ergodic system. They are also good for equidistribution for all totally ergodic systems if and only if the polynomials are rationally independent; this follows easily from a well known equidistribution result of Weyl. If $c_1, \ldots, c_\ell$ are positive distinct non-integers, then it can be shown [9] that the collection of sequences $[n^{c_1}], \ldots, [n^{c_k}]$ is good for seminorm estimates and good for equidistribution for all ergodic systems.

2.2 Main result

We are now ready to state our main result (in applications we are going to use it for $S = [0, 1)$ and $S = ([0, 1) \setminus \mathbb{Q}) \cup \{0\}$.).

Theorem 2.1. *Let S be a susbet of $[0, 1)$ with countable complement in $[0, 1)$. The collection of sequences $a_1, \ldots, a_\ell \colon \mathbb{N} \to \mathbb{Z}$ is jointly ergodic for all systems with spectrum in S if and only if it is good for seminorm estimates and equidistribution for these systems.*

Remarks. • The necessity of the conditions is easy to establish, the interesting part is the sufficiency.

• Theorem 1.1 in [11] uses somewhat weaker assumptions. The stronger assumption we use here allows to simplify the proof in [11].

• Theorem 1.4 in [11] shows that under weaker equidistribution hypothesis, which are satisfied by collections of rationally independent integer polynomials, the rational Kronecker factor controls the limiting behavior of the associated multiple ergodic averages. One can deduce this result from Theorem 2.1 as in [11, Section 5].

In order to facilitate understanding, we are going to first prove Theorem 2.1 for $\ell = 2$ in Section 4 and then explain the necessary changes needed for the proof of the general case in Section 5.

Since for totally ergodic systems $(X, \mathcal{X}, \mu, T)$ we have $\mathrm{Spec}(T) \subset ([0, 1) \setminus \mathbb{Q}) \cup \{0\}$, an immediate consequence of Theorem 2.1 (for $S = ([0, 1) \setminus \mathbb{Q}) \cup \{0\}$) is the following result:

Corollary 2.2. *The collection of sequences $a_1, \ldots, a_\ell \colon \mathbb{N} \to \mathbb{Z}$ is jointly ergodic for all totally ergodic systems if and only if it is good for seminorm estimates for all totally ergodic systems and* (2.2) *holds for all $t_1, \ldots, t_\ell \in [0, 1)$ that are either irrational or zero but not all of them zero.*

This applies to collections of rationally independent polynomials $p_1, \ldots, p_\ell \in \mathbb{Z}[t]$, hence we recover the limit formulas (1.2) and (1.3).

2.3 Applications

Theorem 2.1 can be used to give significantly simpler proofs of results in [3, 9, 13, 16, 20] (the parts that correspond to joint ergodicity properties). But it also gives access to convergence results not previously known. The main reason why Theorem 2.1 is advantageous for these applications, is that it enables us to bypass some difficult and often inaccessible equidistribution results on nilmanifolds that need to be established in order to use the Host-Kra theory of characteristic factors. We record a few instances of these applications below. We remark that in all these cases the most difficult component is to verify the good seminorm property; verifying the needed good equidistribution property is usually a simple matter.

Theorem 2.1 was used in [12] to prove the following joint ergodicity result for sequences given by fractional powers of primes.

Theorem 2.3 ([12]). *Let $c_1, \ldots, c_\ell$ be distinct positive non-integers. Then the collection of sequences $[p_n^{c_1}], \ldots, [p_n^{c_\ell}]$ is jointly ergodic for every ergodic system.*

Previously this was only known for $\ell = 1$ and for $\ell = 2$ it was not even known for nilsystems or weakly mixing systems.

Another very interesting application of Theorem 2.1 was recently obtained by Tsinas [29] who verified a conjecture of the author from [9] (see also [10, Problem 23]).

Theorem 2.4 ([29]). *Let $a_1, \ldots, a_\ell \colon [1, \infty) \to \mathbb{R}$ be functions from a Hardy field*[2] *that have polynomial growth. Then the collection of sequences $[a_1(n)], \ldots, [a_\ell(n)]$ is jointly ergodic for all ergodic systems if whenever $a(t)$ is a non-trivial linear combination of the functions $a_1, \ldots, a_\ell$ we have*

$$\lim_{t \to \infty} \frac{|a(t) - p(t)|}{\log t} = \infty$$

for all polynomials $p \in \mathbb{Z}[t]$.[3]

Previously this was known for $\ell = 1$ (it follows easily from [6]) and for general ℓ partial progress was made in [4, 9, 20, 27].

Theorem 2.1 was also used in [11] in order to address a problem of Bergelson, Moreira, and Richter [4, Conjecture 6.1]. It establishes an extension of the limit formula (1.3) that covers iterates given by polynomials with fractional powers.

Theorem 2.5. *Let $a_1, \ldots, a_\ell \colon \mathbb{R}_+ \to \mathbb{R}$ be linearly independent functions of the form $\sum_{i=1}^{k} \alpha_i t^{c_i}$ where $\alpha_1, \ldots, \alpha_k \in \mathbb{Q}$ and $c_1, \ldots, c_k \in (0, +\infty)$. Then the collection of sequences $[a_1(n)], \ldots, [a_\ell(n)]$ is jointly ergodic for all totally ergodic systems.*

Lastly, Theorem 2.1 was recently extended by Best and Moragues [5] to a large class of countable Abelian group actions, and this extension was subsequently used by Donoso, Koutsogiannis, and Sun [8] to prove joint ergodicity results for commuting transformations with polynomial iterates under some ergodicity assumptions.

3 Background

3.1 Measure preserving systems

A *measure preserving system*, or simply *a system*, is a quadruple $(X, \mathcal{X}, \mu, T)$ where $(X, \mathcal{X}, \mu)$ is a Lebesgue probability space and $T \colon X \to X$ is an invertible, measurable, measure preserving transformation. Throughout, for $n \in \mathbb{N}$ we denote by T^n the composition $T \circ \cdots \circ T$ (n times) and let $T^{-n} := (T^n)^{-1}$ and $T^0 := \mathrm{id}_X$. Also, for $f \in L^2(\mu)$ and $n \in \mathbb{Z}$ we denote by $T^n f$ the function $f \circ T^n$.

[2] This class includes all linear combinations of the functions $t^a (\log t)^b (\log\log t)^c$, $a, b, c \in \mathbb{R}$, and more generally, all functions defined on some half-line $[c, \infty)$ using a finite combination of the symbols $+, -, \times, :$, $\log, \exp$, operating on the real variable t and on real constants.

[3] This condition is close to being necessary, in the sense that if it fails for some non-linear p, then the collection of sequences $a_1, \ldots, a_\ell$ is not going to be jointly ergodic for some ergodic rotation on the ℓ-dimensional torus.

We say that the system $(X, \mathcal{X}, \mu, T)$ is *ergodic* if the only functions $f \in L^2(\mu)$ that satisfy $Tf = f$ are the constant ones. It is *totally ergodic* if $(X, \mathcal{X}, \mu, T^d)$ is ergodic for every $d \in \mathbb{N}$, or equivalently, if the system is ergodic and $\mathrm{Spec}(T) \subset ([0,1) \setminus \mathbb{Q}) \cup \{0\}$.

A function $f \in L^2(\mu)$ is an *eigenfunction* of the system if $Tf = e(\alpha)f$ for some $\alpha \in \mathbb{R}$. We denote with $\mathcal{E}(T)$ the set of all eigenfunctions of the system with unit modulus.

3.2 Gowers-Host-Kra seminorms

Throughout, we use the following notation:

Definition. Let $(X, \mathcal{X}, \mu, T)$ be a system and $f \in L^\infty(\mu)$. If $\underline{n} = (n_1, \ldots, n_s) \in \mathbb{Z}^s$, $\underline{n}' = (n'_1, \ldots, n'_s) \in \mathbb{Z}^s$, $\epsilon = (\epsilon_1, \ldots, \epsilon_s) \in \{0,1\}^s$, and $z \in \mathbb{C}$, we let

(i) $\epsilon \cdot \underline{n} := \epsilon_1 n_1 + \cdots + \epsilon_s n_s$;

(ii) $|\underline{n}| := |n_1| + \cdots + |n_s|$;

(iii) $\mathcal{C}^l z := z$ if l is even and $\mathcal{C}^l z = \overline{z}$ if l is odd;

(iv) $\Delta_n f := T^n f \cdot \overline{f}$, $n \in \mathbb{Z}$;

(v) $\Delta_{\underline{n}} f := \Delta_{n_1} \cdots \Delta_{n_s} f = \prod_{\epsilon \in \{0,1\}^s} \mathcal{C}^{|\epsilon|} T^{\epsilon \cdot \underline{n}} f$.

For instance, we have

$$\Delta_{(n_1,n_2)} f = f \cdot T^{n_1}\overline{f} \cdot T^{n_2}\overline{f} \cdot T^{n_1+n_2} f, \quad n_1, n_2 \in \mathbb{Z}.$$

Given an ergodic system $(X, \mathcal{X}, \mu, T)$ we will make extensive use of the seminorms $\|\!|\cdot|\!\|_s$, $s \in \mathbb{N}$, on $L^\infty(\mu)$, that were introduced in [17]. They are often refereed to as *Gowers-Host-Kra seminorms*, or *uniformity seminorms*, and are defined inductively for $f \in L^\infty(\mu)$ as follows:

$$\|\!|f|\!\|_1 := \Big| \int f \, d\mu \Big|,$$

and for $s \in \mathbb{Z}_+$ we let

$$\|\!|f|\!\|_{s+1}^{2^{s+1}} := \lim_{N\to\infty} \mathbb{E}_{n\in[N]} \|\!|\Delta_n f|\!\|_s^{2^s}. \tag{3.1}$$

For instance, we have

$$\|\!|f|\!\|_2^4 = \lim_{N\to\infty} \mathbb{E}_{n\in[N]} \Big| \int \overline{f} \cdot T^n f \, d\mu \Big|^2.$$

An application of the mean ergodic theorem shows that

$$\|\!|f|\!\|_2^4 = \lim_{N_1\to\infty} \mathbb{E}_{n_1\in[N_1]} \lim_{N_2\to\infty} \mathbb{E}_{n_2\in[N_2]} \int f \cdot T^{n_1}\overline{f} \cdot T^{n_2}\overline{f} \cdot T^{n_1+n_2} f \, d\mu. \tag{3.2}$$

Likewise, by successive applications of the mean ergodic theorem, it can be shown that the limit in (3.1) exists and for $f \in L^\infty(\mu)$ and $s \in \mathbb{Z}_+$ we have that (see [17] or [19, Chapter 8])

$$\|\!|f|\!\|_s^{2^s} = \lim_{N_1\to\infty} \cdots \lim_{N_s\to\infty} \mathbb{E}_{n_1\in[N_1]} \cdots \mathbb{E}_{n_s\in[N_s]} \int \Delta_{(n_1,\ldots,n_s)} f \, d\mu. \tag{3.3}$$

For $s' \in [s]$ it can be shown that we can take any s' of the iterative limits to be simultaneous limits (i.e. average over $[N]^{s'}$ and let $N \to \infty$) without changing the value of the limit. This was originally proved in [17] and for a much simpler proof see [2]. Taking $s' = s$ gives

$$\|f\|_s^{2^s} = \lim_{N\to\infty} \mathbb{E}_{\underline{n}\in[N]^s} \int \Delta_{\underline{n}} f \, d\mu. \tag{3.4}$$

For $s \geq 2$ taking $s' = s-1$ and using the mean ergodic theorem gives

$$\|f\|_s^{2^s} = \lim_{N\to\infty} \mathbb{E}_{\underline{n}\in[N]^{s-1}} \Big| \int \Delta_{\underline{n}} f \, d\mu \Big|^2. \tag{3.5}$$

Lastly, for $s \geq 3$ taking $s' = s-2$ gives

$$\|f\|_s^{2^s} = \lim_{N\to\infty} \mathbb{E}_{\underline{n}\in[N]^{s-2}} \|\Delta_{\underline{n}} f\|_2^4. \tag{3.6}$$

3.3 Soft inverse theorems

Recall that if $(X, \mathcal{X}, \mu, T)$ is a system, with $\mathcal{E}(T)$ we denote the set of its eigenfunctions with unit modulus.

Proposition 3.1. *Let $(X, \mathcal{X}, \mu, T)$ be an ergodic system and $f \in L^\infty(\mu)$ be 1-bounded. Then*

$$\|f\|_2^4 \leq \sup_{\chi\in\mathcal{E}(T)} \Re\Big(\int f \cdot \chi \, d\mu \Big).$$

Proof. Let $\mathcal{K}(T)$ be the closed subspace of $L^2(\mu)$ spanned by all eigenfunctions of the system. It is not hard to prove (see for example [19, Chapter 8, Theorem 1]) that

$$\|f\|_2 = \|\tilde{f}\|_2$$

where $\tilde{f} := \mathbb{E}(f|\mathcal{K}(T))$. Since the system is ergodic and the underlying probability space is Lebesgue, the subspace $\mathcal{K}(T)$ has an orthonormal basis of eigenfunctions of modulus one, say $(\chi_j)_{j\in\mathbb{N}}$. Then $\tilde{f} = \sum_{j=1}^\infty c_j \, \chi_j$ where

$$c_j := \int \tilde{f} \cdot \overline{\chi}_j \, d\mu = \int f \cdot \overline{\chi}_j \, d\mu, \quad j \in \mathbb{N}.$$

We have

$$\|\tilde{f}\|_2^4 = \sum_{j=1}^\infty |c_j|^4 \leq \sup_{j\in\mathbb{N}}(|c_j|^2) \sum_{j=1}^\infty |c_j|^2 = \sup_{j\in\mathbb{N}}(|c_j|^2) \, \|f\|_{L^2(\mu)}^2 \leq \sup_{j\in\mathbb{N}} \Big| \int f \cdot \overline{\chi}_j \, d\mu \Big|,$$

where the first identity follows by orthonormality and direct computation using (3.2), the second identity follows by the Parseval identity, and the last estimate holds since all functions involved are 1-bounded. The result now follows since the set $\mathcal{E}(T)$ is invariant under multiplication by unit modulus constants. □

Proposition 3.2. *Let $(X, \mathcal{X}, \mu, T)$ be an ergodic system and $f \in L^\infty(\mu)$ be such that $\|f\|_{s+2} > 0$ for some $s \in \mathbb{Z}_+$.*

(i) *If $s = 0$, then there exists $\chi \in \mathcal{E}(T)$ such that $\Re\big(\int f \cdot \chi \, d\mu \big) > 0$.*

(ii) *If $s\geq 1$, then there exist $\chi_{\underline{n}}\in \mathcal{E}(T)$, $\underline{n}\in \mathbb{N}^s$, such that*

$$\liminf_{N\to\infty} \mathbb{E}_{\underline{n}\in[N]^s}\Re\Big(\int \Delta_{\underline{n}}f\cdot \chi_{\underline{n}}\, d\mu\Big)>0.$$

Proof. If $s=0$, then the conclusion follows immediately from Proposition 3.1.

Suppose that $s\geq 1$. By (3.6) we have that

$$\lim_{N\to\infty} \mathbb{E}_{\underline{n}\in[N]^s}|||\Delta_{\underline{n}}f|||_2^4>0.$$

Using Proposition 3.1 we deduce that

$$\liminf_{N\to\infty} \mathbb{E}_{\underline{n}\in[N]^s}\sup_{\chi\in\mathcal{E}(T)}\Re\Big(\int \Delta_{\underline{n}}f\cdot \chi\, d\mu\Big)>0.$$

This immediately implies the asserted estimate. □

3.4 Gowers-Cauchy-Schwarz estimates

We will use the following variant of the so called Gowers-Cauchy-Schwarz inequality:

Lemma 3.3. *Let $(X,\mathcal{X},\mu,T)$ be a system, for $s\in\mathbb{N}$ let $f_\epsilon\in L^\infty(\mu)$, $\epsilon\in\{0,1\}^s$, be 1-bounded functions, and $g_{\underline{n}}\in L^\infty(\mu)$, $\underline{n}\in\mathbb{N}^s$. Let also $\underline{1}:=(1,\ldots,1)$. Then for every $N\in\mathbb{N}$ we have*

$$\Big|\mathbb{E}_{\underline{n}\in[N]^s}\int\prod_{\epsilon\in\{0,1\}^s}T^{\epsilon\cdot\underline{n}}f_\epsilon\cdot g_{\underline{n}}\, d\mu\Big|^{2^s}\leq \mathbb{E}_{\underline{n},\underline{n}'\in[N]^s}\int \Delta_{\underline{n}-\underline{n}'}f_{\underline{1}}\cdot T^{-|\underline{n}|}g_{\underline{n},\underline{n}'}\, d\mu,$$

where for every $\underline{n},\underline{n}'\in\mathbb{N}^s$ the function $g_{\underline{n},\underline{n}'}$ is equal to a product of 2^s functions that belong to the set $\{g_{\underline{n}},\overline{g}_{\underline{n}},\underline{n}\in\mathbb{N}^s\}$.

Proof. For notational simplicity we give the details only for $s=2$. The general case can be proved in a similar manner by successively applying the Cauchy-Schwarz inequality with respect to the variables $n_s,\ldots,n_1$, exactly as we do below for $s=2$. We have that

$$\Big|\mathbb{E}_{n_1,n_2\in[N]}\int f_0\cdot T^{n_1}f_1\cdot T^{n_2}f_2\cdot T^{n_1+n_2}f_3\cdot g_{n_1,n_2}\, d\mu\Big|^2$$

is bounded by (we use that f_0, f_1 are 1-bounded)

$$\mathbb{E}_{n_1\in[N]}\int\Big|\mathbb{E}_{n_2\in[N]}T^{n_2}f_2\cdot T^{n_1+n_2}f_3\cdot g_{n_1,n_2}\Big|^2\, d\mu.$$

After expanding the square we find that this expression is equal to

$$\mathbb{E}_{n_1\in[N]}\int \mathbb{E}_{n_2,n_2'\in[N]}T^{n_2}f_2\cdot T^{n_2'}\overline{f}_2\cdot T^{n_1+n_2}f_3\cdot T^{n_1+n_2'}\overline{f}_3\cdot g_{n_1,n_2}\cdot \overline{g_{n_1,n_2'}}\, d\mu.$$

After composing with T^{-n_2}, exchanging $\mathbb{E}_{n_1\in[N]}$ with $\mathbb{E}_{n_2,n_2'\in[N]}$, using the Cauchy-Schwarz inequality, and that f_2 is 1-bounded, we get that the square of the last expression is bounded by

$$\mathbb{E}_{n_2,n_2'\in[N]}\int\Big|\mathbb{E}_{n_1\in[N]}T^{n_1}f_3\cdot T^{n_1+n_2'-n_2}\overline{f}_3\cdot T^{-n_2}(g_{n_1,n_2}\cdot\overline{g}_{n_1,n_2'})\Big|^2\, d\mu.$$

As before, we expand the square, and compose with T^{-n_1}. We arrive at the expression

$$\mathbb{E}_{n_1,n_2,n_1',n_2'\in[N]}$$
$$\int f_3 \cdot T^{n_1'-n_1}\overline{f_3} \cdot T^{n_2'-n_2}\overline{f_3} \cdot T^{n_1'+n_2'-n_1-n_2} f_3 \cdot T^{-n_1-n_2}(g_{n_1,n_2} \cdot \overline{g}_{n_1,n_2'} \cdot \overline{g}_{n_1',n_2} \cdot g_{n_1',n_2'})\, d\mu,$$

which is equal to the right hand side of the asserted estimate when $s = 2$ (for $\underline{n} := (n_1, n_2)$, $\underline{n}' := (n_1', n_2')$). Combining the previous two estimates gives the asserted bound for $s = 2$. □

4 Proof of the main result for $\ell = 2$

The goal of this section is to give a proof for the sufficiency of the conditions in Theorem 2.1 for $\ell = 2$ (the necessity is simple). It suffices to prove the following:

Theorem 4.1. *Let S be a susbet of $[0, 1)$ with countable complement in $[0, 1)$. Suppose that the sequences $a, b\colon \mathbb{N} \to \mathbb{Z}$ are good for equidistribution on S and seminorm estimates for the system $(X, \mathcal{X}, \mu, T)$ with $\operatorname{Spec}(T) \subset S$. Then for all $f, g \in L^\infty(\mu)$ we have*

$$\lim_{N\to\infty} \mathbb{E}_{n\in[N]}\, T^{a(n)} f \cdot T^{b(n)} g = \int f\, d\mu \cdot \int g\, d\mu \tag{4.1}$$

in $L^2(\mu)$.

The proof of Theorem 2.1 for general ℓ is similar to the case $\ell = 2$ but involves an additional induction and is notationally more complicated; we describe the modifications needed to get the more general statement in Section 5.

4.1 Preparation

In order to ease the exposition of the proof of Theorem 4.1 we use this subsection to gather some preparatory results. We are going to complete the proof of Theorem 4.1 in Section 4.2.

4.1.1 The case where g is an eigenfunction

We are going to make essential use of the good equidistribution assumption for the sequences $a, b\colon \mathbb{N} \to \mathbb{Z}$ to prove the next result.

Proposition 4.2. *Theorem 4.1 holds if g is an eigenfunction of the system.*

Proof. If f is constant, then the conclusion easily follows from our equidistribution assumption. Thus, it suffices to show that if $\int f\, d\mu = 0$ and $\chi \in \mathcal{E}(T)$, then

$$\lim_{N\to\infty} \mathbb{E}_{n\in[N]}\, T^{a(n)} f \cdot T^{b(n)} \chi = 0$$

in $L^2(\mu)$.

Suppose that the eigenvalue of χ is $e(\alpha)$ for some $\alpha \in \mathrm{Spec}(T)$. Then $T^{b(n)}\chi = e(b(n)\alpha)\,\chi$, so it suffices to show that

$$\lim_{N\to\infty} \mathbb{E}_{n\in[N]}\, c_n\, T^{a(n)} f = 0 \tag{4.2}$$

in $L^2(\mu)$ where $c_n := e(b(n)\alpha)$, $n \in \mathbb{N}$. To this end, we invoke the theorem of Herglotz (see for example [21, Section 7.6]) for the positive definite sequence $a(n) := \int \overline{f} \cdot T^n f\, d\mu$, $n \in \mathbb{Z}$. It gives that there exists a positive bounded measure σ on $\mathbb{T}$ (thought of as $[0,1)$) such that

$$\int \overline{f} \cdot T^n f\, d\mu = \int e(nt)\, d\sigma(t), \quad n \in \mathbb{Z}. \tag{4.3}$$

Note that σ does not have a point mass on 0 because f has integral 0, or on any other number on the complement of $\mathrm{Spec}(T)$ (we leave these standard facts as an exercise for the reader). A simple computation that uses (4.3) shows that

$$\left\|\mathbb{E}_{n\in[N]}\, c_n\, T^{a(n)} f\right\|_{L^2(\mu)} = \left\|\mathbb{E}_{n\in[N]}\, c_n\, e(a(n)t)\right\|_{L^2(\sigma)}, \quad N \in \mathbb{N}.$$

Using this identity, the bounded convergence theorem, and the fact that the bounded measure σ does not have point masses on 0 and on the countable set $[0,1) \setminus S$ (since it is contained on the complement of $\mathrm{Spec}(T)$), we get that (4.2) would follow if we show that for every non-zero $t \in S$ we have

$$\lim_{N\to\infty} \mathbb{E}_{n\in[N]}\, e(a(n)t + b(n)\alpha) = 0.$$

Since $\alpha \in \mathrm{Spec}(T) \subset S$, this follows from our assumption that the pair of sequences $a, b\colon \mathbb{N} \to \mathbb{Z}$ is good for equidistribution on S. □

4.1.2 Positivity for g implies positivity for an averaged function $\tilde{g}$

Our next goal is to show that if the positivity property (4.4) below holds, then it also holds when we replace g with an averaged function $\tilde{g}$. This is a simple but crucial observation because uniformity properties of $\tilde{g}$ are easier to analyse than those of g.

Proposition 4.3. *Let $(X, \mathcal{X}, \mu, T)$ be a system and $f, g \in L^\infty(\mu)$ be such that*

$$\limsup_{N\to\infty} \left\|\mathbb{E}_{n\in[N]}\, T^{a(n)} f \cdot T^{b(n)} g\right\|_{L^2(\mu)} > 0. \tag{4.4}$$

Then there exist $N_k \to \infty$ and 1-bounded $g_k \in L^\infty(\mu)$, $k \in \mathbb{N}$, such that for

$$\tilde{g} := \lim_{k\to\infty} \mathbb{E}_{n\in[N_k]}\, T^{-b(n)} g_k \cdot T^{a(n)-b(n)}\overline{f}, \tag{4.5}$$

where the limit is a weak limit (note that then $\tilde{g} \in L^\infty(\mu)$), we have

$$\limsup_{N\to\infty} \left\|\mathbb{E}_{n\in[N]}\, T^{a(n)} f \cdot T^{b(n)} \tilde{g}\right\|_{L^2(\mu)} > 0. \tag{4.6}$$

Proof. We can assume that both f and g are 1-bounded. For fixed $f \in L^\infty(\mu)$ we let $\mathcal{C} = \mathcal{C}(f)$ be the $L^2(\mu)$ closure of all linear combinations of all subsequential weak-limits of sequences of the form $\mathbb{E}_{n\in[N]}\, T^{-b(n)}g_N \cdot T^{-b(n)+a(n)}\overline{f}$, where $g_N \in L^\infty(\mu)$, $N\in\mathbb{N}$, are 1-bounded functions.

We first claim that if g is orthogonal to the subspace $\mathcal{C}$, then

$$\lim_{N\to\infty} \mathbb{E}_{n\in[N]}\, T^{a(n)}f\cdot T^{b(n)}g = 0$$

in $L^2(\mu)$. Indeed, if this is not the case, then there exist $a>0$ and $N_k\to\infty$ such that

$$\Big\|\mathbb{E}_{n\in[N_k]}\, T^{a(n)}f\cdot T^{b(n)}g\Big\|_{L^2(\mu)} \geq a, \quad k\in\mathbb{N}.$$

If we define the 1-bounded functions $g_k := \mathbb{E}_{n\in[N_k]}\, T^{a(n)}f\cdot T^{b(n)}g$, $k\in\mathbb{N}$, we deduce that

$$\mathbb{E}_{n\in[N_k]} \int \overline{g}_k\cdot T^{a(n)}f\cdot T^{b(n)}g\, d\mu \geq a^2, \quad k\in\mathbb{N}. \tag{4.7}$$

By passing to a subsequence, we can assume that the sequence of 1-bounded functions $\mathbb{E}_{n\in[N_k]}\, T^{-b(n)}g_k\cdot T^{a(n)-b(n)}\overline{f}$, $k\in\mathbb{N}$, converges in the weak topology of $L^2(\mu)$ to a function $h\in\mathcal{C}$. Then composing with $T^{-b(n)}$ in (4.7) we deduce that $\int g\cdot\overline{h}\, d\mu\neq 0$, contradicting our assumption that g is orthogonal to the subspace $\mathcal{C}$. This proves our claim.

From the previous claim we conclude that

$$\lim_{N\to\infty} \mathbb{E}_{n\in[N]}\, T^{a(n)}f\cdot T^{b(n)}(g-\mathbb{E}(g|\mathcal{C})) = 0$$

in $L^2(\mu)$, where $\mathbb{E}(g|\mathcal{C})$ denotes the orthogonal projection of g onto the closed subspace $\mathcal{C}$. Hence, if (4.4) holds, then

$$\limsup_{N\to\infty} \Big\|\mathbb{E}_{n\in[N]}\, T^{a(n)}f\cdot T^{b(n)}\mathbb{E}(g|\mathcal{C})\Big\|_{L^2(\mu)} > 0.$$

Using the definition of $\mathcal{C}$ and an approximation argument, we get that there exist $N_k\to\infty$ and 1-bounded functions $g_k\in L^\infty(\mu)$, $k\in\mathbb{N}$, such that for $\tilde{g}$ as in (4.5) we have that (4.6) holds. Lastly, since f and g_k, $k\in\mathbb{N}$, are 1-bounded functions, the same holds for $\tilde{g}$. This completes the proof. □

4.1.3 Seminorms of averaged functions

Our next goal is to use Lemma 3.3 in order to show that if the uniformity seminorm of an average of functions is positive, then some positiveness property holds for iterated differences of the individual functions. Note that we do not impose any assumptions on the sequences $a,b\colon\mathbb{N}\to\mathbb{Z}$ here.

Proposition 4.4. *Let $(X,\mathcal{X},\mu,T)$ be an ergodic system, and $\tilde{g}\in L^\infty(\mu)$ be as in* (4.5) *and satisfy $\|\tilde{g}\|_{s+2}>0$ for some $s\in\mathbb{Z}_+$.*

(i) *If $s=0$, then there exists $\chi\in\mathcal{E}(T)$ such that*

$$\limsup_{k\to\infty} \mathbb{E}_{n\in[N_k]}\, \Re\Big(\int g_k\cdot T^{a(n)}\overline{f}\cdot T^{b(n)}\chi\, d\mu\Big) > 0.$$

(ii) *If $s \geq 1$, then there exist $\chi_{\underline{n},\underline{n}'} \in \mathcal{E}(T)$, $\underline{n}, \underline{n}' \in \mathbb{N}^s$, such that*

$$\liminf_{N\to\infty} \mathbb{E}_{\underline{n},\underline{n}'\in[N]^s} \limsup_{k\to\infty} \mathbb{E}_{n\in[N_k]} \Re\Big(\int (\Delta_{\underline{n}-\underline{n}'} g_k)\cdot T^{a(n)}(\Delta_{\underline{n}-\underline{n}'}\overline{f})\cdot T^{b(n)}\chi_{\underline{n},\underline{n}'}\, d\mu\Big) > 0.$$

Remark 1. The key point is that positivity properties of expressions involving $\Delta_{\underline{n}}\tilde{g}$, $\underline{n} \in \mathbb{N}^{s+2}$, imply positivity properties of expressions involving $\Delta_{\underline{n}} f$, $\underline{n} \in \mathbb{N}^s$.

Proof. Suppose that $s \geq 1$, the argument is similar if $s = 0$. Proposition 3.2 gives that there exist $\chi_{\underline{n}} \in \mathcal{E}(T)$, $\underline{n} \in \mathbb{N}^s$, such that

$$\liminf_{N\to\infty} \mathbb{E}_{\underline{n}\in[N]^s} \Re\Big(\int \Delta_{\underline{n}}\tilde{g} \cdot \chi_{\underline{n}}\, d\mu\Big) > 0.$$

Since $\Delta_{\underline{n}}\tilde{g} = \prod_{\epsilon\in\{0,1\}^s} \mathcal{C}^{|\epsilon|} T^{\epsilon\cdot\underline{n}}\tilde{g}$, $\underline{n} \in \mathbb{N}^s$, and $\tilde{g} = \lim_{k\to\infty} \mathbb{E}_{n\in[N_k]} f_{k,n}$ (the limit is a weak limit) where

$$f_{k,n} := T^{-b(n)} g_k \cdot T^{a(n)-b(n)}\overline{f}, \quad k, n \in \mathbb{N}, \tag{4.8}$$

we deduce that

$$\liminf_{N\to\infty} \lim_{k\to\infty} \mathbb{E}_{n\in[N_k]} \Re\Big(\mathbb{E}_{\underline{n}\in[N]^s} \int \prod_{\epsilon\in\{0,1\}^s\setminus\{\underline{1}\}} \mathcal{C}^{|\epsilon|} T^{\epsilon\cdot\underline{n}}\tilde{g} \cdot T^{n_1+\cdots+n_s} f_{k,n} \cdot \chi_{\underline{n}}\, d\mu\Big) > 0.$$

For fixed $k, n \in \mathbb{N}$, we apply Lemma 3.3 with $f_{\underline{1}} := f_{k,n}$, $f_\epsilon := \mathcal{C}^{|\epsilon|} f$ for $\epsilon \in \{0,1\}^s \setminus \underline{1}$, and $g_{\underline{n}} := \chi_{\underline{n}}$, $\underline{n} \in \mathbb{N}^s$, and deduce that

$$\liminf_{N\to\infty} \limsup_{k\to\infty} \mathbb{E}_{n\in[N_k]} \mathbb{E}_{\underline{n},\underline{n}'\in[N]^s} \int \Delta_{\underline{n}-\underline{n}'} f_{k,n} \cdot \chi_{\underline{n},\underline{n}'}\, d\mu > 0$$

for some $\chi_{\underline{n},\underline{n}'} \in \mathcal{E}(T)$, $\underline{n}, \underline{n}' \in \mathbb{N}^s$ (we used that $\mathcal{E}(T)$ is closed under products and composition with iterates of T). Note that $\Delta_{\underline{n}}(w \cdot z) = \Delta_{\underline{n}}(w) \cdot \Delta_{\underline{n}}(z)$ and $\Delta_{\underline{n}}(T^k w) = T^k \Delta_{\underline{n}}(w)$ for all $w, z \in L^\infty(\mu)$ and $k \in \mathbb{N}$, $\underline{n} \in \mathbb{Z}^s$. Using this, equation (4.8), and keeping in mind that the limsup of a sum is at most the sum of the limsups, the asserted estimate follows from the last one after composing each function inside the integral with $T^{b(n)}$. $\square$

4.2 Proof of Theorem 4.1

We are now ready to prove Theorem 4.1.

4.2.1 Reduction to a degree lowering property

Since the sequences $a, b \colon \mathbb{N} \to \mathbb{Z}$ are good for seminorms estimates for the system $(X, \mathcal{X}, \mu, T)$, there exists $s \in \mathbb{Z}_+$ such that the seminorms $\|\!|\cdot\|\!|_{s+2}$ control the averages (4.1), in the sense that if $f, g \in L^\infty(\mu)$ are such that $\|\!| f\|\!|_{s+2} = 0$ or $\|\!| g\|\!|_{s+2} = 0$, then

$$\lim_{N\to\infty} \mathbb{E}_{n\in[N]} T^{a(n)} f \cdot T^{b(n)} g = 0 \tag{4.9}$$

in $L^2(\mu)$. Our goal is to show that a similar property holds with $s - 1$ in place of s. Namely, using terminology from [26], we are going to establish the following *"degree lowering property"*:

Proposition 4.5. *Let S be a susbet of $[0,1)$ with countable complement in $[0,1)$. Let $a, b\colon \mathbb{N} \to \mathbb{Z}$ be good for equidistribution for S and $(X, \mathcal{X}, \mu, T)$ be an ergodic system with spectrum in S. If for some $s \in \mathbb{Z}_+$ the seminorms $\|\!|\cdot|\!\|_{s+2}$ control the averages* (4.1)*, then also the seminorms $\|\!|\cdot|\!\|_{s+1}$ control the averages* (4.1).

This "degree lowering property" is the heart of the proof of Theorem 4.1. Iterating this property $s+1$ times we deduce that the seminorms $\|\!|\cdot|\!\|_1$ control the averages (4.1). Since $\|\!|f|\!\|_1 = |\int f\, d\mu|$, this proves Theorem 4.1.[4] So we get the following:

Proposition 4.6. *In order to verify Theorem 4.1 it suffices to verify Proposition 4.5.*

4.2.2 Proof of Proposition 4.5

We work under the assumption of Proposition 4.5 and our aim is to show that if $f \in L^\infty(\mu)$ satisfies $\|\!|f|\!\|_{s+1} = 0$, then (4.9) holds (similarly we show that if $g \in L^\infty(\mu)$ satisfies $\|\!|g|\!\|_{s+1} = 0$, then (4.9) holds). Equivalently, it suffices to show that if (4.4) holds, then $\|\!|f|\!\|_{s+1} > 0$.

Suppose that $s \geq 1$, the argument is similar if $s = 0$ (in this case the conclusion is that $\int f\, d\mu \neq 0$). Using (4.4) and Proposition 4.3, we deduce that

$$\limsup_{N\to\infty} \left\| \mathbb{E}_{n\in[N]}\, T^{a(n)} f \cdot T^{b(n)} \tilde{g} \right\|_{L^2(\mu)} > 0,$$

where

$$\tilde{g} := \lim_{k\to\infty} \mathbb{E}_{n\in[N_k]}\, T^{-b(n)} g_k \cdot T^{a(n)-b(n)} \overline{f}, \tag{4.10}$$

for some sequence of integers $N_k \to \infty$ and 1-bounded functions $g_k \in L^\infty(\mu)$, $k \in \mathbb{N}$, where the limit is a weak limit. Since, by assumption, the seminorms $\|\!|\cdot|\!\|_{s+2}$ control the averages in (4.9) we get that

$$\|\!|\tilde{g}|\!\|_{s+2} > 0. \tag{4.11}$$

Using Proposition 4.4 we deduce that

$$\liminf_{N\to\infty} \mathbb{E}_{\underline{n},\underline{n}'\in[N]^s} \limsup_{k\to\infty} \mathbb{E}_{n\in[N_k]} \Re\Big(\int (\Delta_{\underline{n}-\underline{n}'} g_k) \cdot T^{a(n)}(\Delta_{\underline{n}-\underline{n}'} \overline{f}) \cdot T^{b(n)} \chi_{\underline{n},\underline{n}'}\, d\mu \Big) > 0$$

for some $\chi_{\underline{n},\underline{n}'} \in \mathcal{E}(T)$, $\underline{n}, \underline{n}' \in \mathbb{N}^s$. Using the Cauchy-Schwarz inequality we get

$$\liminf_{N\to\infty} \mathbb{E}_{\underline{n},\underline{n}'\in[N]^s} \limsup_{k\to\infty} \left\| \mathbb{E}_{n\in[N_k]}\, T^{a(n)}(\Delta_{\underline{n}-\underline{n}'} \overline{f}) \cdot T^{b(n)} \chi_{\underline{n},\underline{n}'} \right\|_{L^2(\mu)} > 0.$$

The advantage now is that since $\chi_{\underline{n},\underline{n}'} \in \mathcal{E}(T)$, $\underline{n}, \underline{n}' \in \mathbb{N}^d$, the average over n is much simpler to analyse than the original one in Theorem 4.1. In fact, using Proposition 4.2 we get that it converges in $L^2(\mu)$ to the product of the integrals of the individual functions. We deduce that

$$\liminf_{N\to\infty} \mathbb{E}_{\underline{n},\underline{n}'\in[N]^s} \left| \int \Delta_{\underline{n}-\underline{n}'} f\, d\mu \cdot \int \chi_{\underline{n},\underline{n}'}\, d\mu \right| > 0$$

and as a consequence

$$\liminf_{N\to\infty} \mathbb{E}_{\underline{n},\underline{n}'\in[N]^s} \left| \int \Delta_{\underline{n}-\underline{n}'} f\, d\mu \right|^2 > 0.$$

[4]We use here that Theorem 4.1 holds if f or g is constant, which follows from Proposition 4.2.

Since $|\int \Delta_{\underline{n}} f\, d\mu|$ remains the same if we change the sign of some of the coordinates of $\underline{n}$, we deduce using a simple computation that

$$\liminf_{N\to\infty} \mathbb{E}_{\underline{n}\in[N]^s} \prod_{j=1}^{s} \big(1 - \frac{n_j}{N}\big) \cdot \Big| \int \Delta_{\underline{n}} f\, d\mu \Big|^2 > 0.$$

It follows that

$$\lim_{N\to\infty} \mathbb{E}_{\underline{n}\in[N]^s} \Big| \int \Delta_{\underline{n}} f\, d\mu \Big|^2 > 0$$

(the limit exists by (3.5)). Hence, by (3.5) we have that

$$|\!|\!| f |\!|\!|_{s+1} > 0$$

as required. This concludes the proof of Proposition 4.5 and by Proposition 4.6 the proof of Theorem 4.1.

5 Proof of the main result for general ℓ

We now give a summary of the proof of Theorem 2.1 for general ℓ, the reader should find it easy to fill in the missing details.

Let S be a susbet of $[0,1)$ with countable complement in $[0,1)$. Suppose that collection of sequences $a_1, \ldots, a_\ell \colon \mathbb{N} \to \mathbb{Z}$ is good for seminorm estimates and equidistribution for the ergodic system $(X, \mathcal{X}, \mu, T)$ with specrum in S. Our goal is to show that for all $f_1, \ldots, f_\ell \in L^\infty(\mu)$ we have

$$\lim_{N\to\infty} \mathbb{E}_{n\in[N]}\, T^{a_1(n)} f_1 \cdot \ldots \cdot T^{a_\ell(n)} f_\ell = \int f_1\, d\mu \cdot \ldots \cdot \int f_\ell\, d\mu \quad \text{in} \quad L^2(\mu). \tag{5.1}$$

Our proof will deviate slightly from the argument given in the case $\ell = 2$, because a statement analogous to Proposition 4.2 cannot be proved directly when only one of the functions is in $\mathcal{E}(T)$ (the theorem of Herglotz is no longer applicable). As a substitute for this we are going to use an induction that we describe next.

We consider $\ell \geq 2$ fixed and we are going to show the following property by finite induction on $m \in \{1, \ldots, \ell\}$:

(P_m) If $f_j \in \mathcal{E}(T)$ for at least $\ell - m$ values of $j \in [\ell]$, then (5.1) holds.

If we show this, then taking $m = \ell$ gives that (5.1) holds for all functions $f_1, \ldots, f_\ell \in L^\infty(\mu)$ (and as a consequence Theorem 2.1 holds).

For $m = 1$ we can show that (P_1) holds as in Proposition 4.2 using the good equidistribution assumption of the sequences $a_1, \ldots, a_\ell$ and the theorem of Herglotz.

Suppose now that property (P_{m-1}) holds for some $m \in \{2, \ldots \ell\}$. We are going to show that property (P_m) holds. To this end, we assume, without loss of generality, that $f_j \in \mathcal{E}(T)$ for $j = m+1, \ldots, \ell$, and we are going to show that (5.1) holds by employing a degree lowering argument, similar to the one we used in the previous section. More precisely, our plan is to show that if for some $s \in \mathbb{Z}_+$ the seminorms $|\!|\!| \cdot |\!|\!|_{s+2}$ control the averages in (5.1), in the sense that if $|\!|\!| f_i |\!|\!|_{s+2} = 0$ for some $i \in \{1, \ldots, \ell\}$ and $f_j \in \mathcal{E}(T)$ for $j = m+1, \ldots, \ell$, we have that the averages in (5.1) converge to 0, then the seminorms

$\|\!| \cdot |\!\|_{s+1}$ also control these averages. Since, by our good seminorm assumption, the seminorms $\|\!| \cdot |\!\|_{s+2}$ control the averages (5.1) for some $s \in \mathbb{N}$, iterating this degree lowering property $s+1$ times we deduce that the seminorms $\|\!| \cdot |\!\|_1$ control the averages in (5.1), and this easily implies that property (P_m) holds.

So suppose that

$$\limsup_{N\to\infty} \Big\| \mathbb{E}_{n\in[N]}\, T^{a_1(n)} f_1 \cdot \ldots \cdot T^{a_\ell(n)} f_\ell \Big\|_{L^2(\mu)} > 0 \tag{5.2}$$

for some 1-bounded functions $f_1, \ldots, f_\ell \in L^\infty(\mu)$ with $f_j \in \mathcal{E}(T)$ for $j = m+1, \ldots, \ell$. Then arguing as in the proof of Proposition 4.3 we get that (5.2) continues to hold if in place of the function f_1 we use the function $\tilde{f}_1$ defined by

$$\tilde{f}_1 := \lim_{k\to\infty} \mathbb{E}_{n\in[N_k]}\, T^{-a_1(n)} g_k \cdot \prod_{j=2}^{\ell} T^{a_j(n)-a_1(n)} \overline{f}_j, \tag{5.3}$$

for some $N_k \to \infty$ and 1-bounded functions $g_k \in L^\infty(\mu)$, $k \in \mathbb{N}$, where the limit is a weak limit (note that then $\tilde{f}_1 \in L^\infty(\mu)$). Since, by our assumption, the seminorms $\|\!|\cdot|\!\|_{s+2}$ control the averages (5.1) we deduce that $\|\!| \tilde{f}_1 |\!\|_{s+2} > 0$. As in the proof of Proposition 4.4 we get for $s \geq 1$ that there exist $\chi_{\underline{n},\underline{n}'} \in \mathcal{E}(T)$, $\underline{n}, \underline{n}' \in \mathbb{N}^s$, such that

$$\liminf_{N\to\infty} \mathbb{E}_{\underline{n},\underline{n}'\in[N]^s} \limsup_{k\to\infty} \mathbb{E}_{n\in[N_k]} \Re\Big(\int (\Delta_{\underline{n}-\underline{n}'} g_k)\cdot T^{a_1(n)} \chi_{\underline{n},\underline{n}'} \cdot \prod_{j=2}^{\ell} T^{a_j(n)} (\Delta_{\underline{n}-\underline{n}'} \overline{f}_j)\, d\mu \Big) > 0,$$

and a somewhat simpler statement for $s = 0$ that can be dealt in a similar fashion. The advantage now is that since for all $\underline{n}, \underline{n}' \in \mathbb{N}^s$ we have $\chi_{\underline{n},\underline{n}'} \in \mathcal{E}(T)$ and $\Delta_{\underline{n}-\underline{n}'} \overline{f}_j \in \mathcal{E}(T)$ for $j = m+1, \ldots, \ell$, property (P_{m-1}) applies and gives that the average over the variable n converges in $L^2(\mu)$ to the product of the integrals of the corresponding functions. We then deduce as in the proof of Proposition 4.6 that

$$\|\!| f_j |\!\|_{s+1} > 0 \quad \text{for} \quad j = 2, \ldots \ell.$$

Furthermore, since $l \geq 2$ we can apply the same argument for the second position instead of the first and deduce in a similar fashion that $\|\!| f_1 |\!\|_{s+1} > 0$. We conclude that the seminorms $\|\!| \cdot |\!\|_{s+1}$ control the averages (5.1). This shows that property (P_m) holds and concludes the proof of the induction and the proof of Theorem 2.1.

References

[1] V. Bergelson, A. Leibman. Polynomial extensions of van der Waerden's and Szemerédi's theorems. *J. Amer. Math. Soc.* **9** (1996), 725–753.

[2] V. Bergelson, A. Leibman. Cubic averages and large intersections. Recent trends in ergodic theory and dynamical systems. *Contemporary Mathematics* **631** (2015), 5–19.

[3] V. Bergelson, J. Moreira, F. Richter. Single and multiple recurrence along non-polynomial sequences. *Adv. Math.* **368** (2020), 107–146.

[4] V. Bergelson, J. Moreira, F. Richter. Multiple ergodic averages along functions from a Hardy field: convergence, recurrence and combinatorial applications. Preprint 2020, arXiv:2006.03558v1

[5] A. Best, A. F. Moragues. Polynomial ergodic averages for certain countable ring actions. To appear in *Discrete Contin. Dyn. Syst.*, `arXiv:2105.04008`

[6] M. Boshernitzan. Uniform distribution and Hardy fields. *J. Analyse Math.* **62** (1994), 225–240.

[7] J-P. Conze, E. Lesigne. Théorèmes ergodiques pour des mesures diagonales. *Bull. Soc. Math. France* **112** (1984), no. 2, 143–175.

[8] S. Donoso, A. Koutsogiannis, W. Sun. Seminorms for multiple averages along polynomials and applications to joint ergodicity. To appear in *J. Analyse Math.*, `arXiv:1902.10237`

[9] N. Frantzikinakis. Multiple recurrence and convergence for Hardy sequences of polynomial. *J. Analyse Math.* **112** (2010), 79–135.

[10] N. Frantzikinakis. Some open problems on multiple ergodic averages. *Bull. Hellenic Math. Soc.* **60** (2016), 41–90.

[11] N. Frantzikinakis. Joint ergodicity of sequences. Preprint 2021, `arXiv:2102.09967`

[12] N. Frantzikinakis. Joint ergodicity of fractional powers of primes. *Forum of Mathematics, Sigma* **10** (2022), e30.

[13] N. Frantzikinakis, B. Kra. Polynomial averages converge to the product of integrals. *Isr. J. Math.* **148** (2005), 267–276.

[14] H. Furstenberg. Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J. Analyse Math.* **31** (1977), 204–256.

[15] H. Furstenberg. *Recurrence in ergodic theory and combinatorial number theory.* Princeton University Press, Princeton, 1981.

[16] H. Furstenberg, B. Weiss. A mean ergodic theorem for $(1/N)\sum_{n=1}^N f(T^n x) \quad g(T^{n^2}x)$. Convergence in ergodic theory and probability (Columbus, OH, 1993), Ohio State Univ. Math. Res. Inst. Publ., **5**, de Gruyter, Berlin, (1996), 193–227.

[17] B. Host, B. Kra. Non-conventional ergodic averages and nilmanifolds. *Ann. of Math.* **161** (2005), 397–488.

[18] B. Host, B. Kra. Convergence of polynomial ergodic averages. *Isr. J. Math.* **149** (2005), 1–19.

[19] B. Host, B. Kra. *Nilpotent Structures in Ergodic Theory.* Mathematical Surveys and Monographs, vol. 236. American Mathematical Society, Providence, RI, 2018.

[20] D. Karageorgos, A. Koutsogiannis. Integer part independent polynomial averages and applications along primes. *Studia Mathematica* **249** (2019), 233–257.

[21] Y. Katznelson. *An introduction to harmonic analysis.* Third edition. Cambridge Mathematical Library, Cambridge University Press, 2004.

[22] A. Leibman. Pointwise convergence of ergodic averages for polynomial sequences of rotations of a nilmanifold. *Ergodic Theory Dynam. Systems* **25** (2005), no. 1, 201–213.

[23] A. Leibman. Convergence of multiple ergodic averages along polynomials of several variables. *Isr. J. Math.* **146** (2005), 303–316.

[24] S. Peluse. On the polynomial Szemerédi theorem in finite fields. *Duke Math. J.* **168** (2019), 749–774.

[25] S. Peluse. Bounds for sets with no polynomial progressions. *Forum of Mathematics, Pi*, **8** (2020), e16.

[26] S. Peluse, S. Prendiville. Quantitative bounds in the non-linear Roth theorem. Preprint 2019, `arXiv:1903.02592`

[27] F. Richter. Uniform distribution in nilmanifolds along functions from a Hardy field. To appear in *J. Analyse Math.*, `arXiv:2006.02028`

[28] E. Szemerédi. On sets of integers containing no k elements in arithmetic progression. *Acta Arith.* **27** (1975), 299–345.

[29] K. Tsinas. Joint ergodicity of Hardy field sequences. To appear in *Trans. Amer. Math. Soc.*, `arXiv:2109.07941`

[30] T. Ziegler. A nonconventional ergodic theorem for a nilsystem. *Ergodic Theory Dynam. Systems* **25** (2005), no. 4, 1357–1370.

[31] T. Ziegler. Universal characteristic factors and Furstenberg averages. *J. Amer. Math. Soc.* **20** (2007), 53–97.

Nikos Frantzikinakis: University of Crete, Department of mathematics and applied mathematics, Voutes University Campus, Heraklion 71003, Greece.

E-mail: frantzikinakis@gmail.com

A sharp upper bound for the Hausdorff dimension of the set of exceptional points for the strong density theorem

Panagiotis Georgopoulos and Constantinos Gryllakis

Abstract

Given an $E \subseteq \mathbb{R}^m$, Lebesgue measurable, we construct a real function $\psi : \mathbb{R}^+ \to \mathbb{R}^+$ (depending on E) increasing, with $\lim\limits_{t \to 0^+} \psi(t) = 0$ such that

$$\lim_{\substack{x \in R \\ d(R) \to 0}} \frac{|R \cap E^c|}{|R| \cdot \psi(d(R))} = 0 \qquad \text{for a.e. } x \in E$$

(where R is an interval in $\mathbb{R}^m$ and d stands for the diameter). This gives a new constructive proof of a problem posed by S. J. Taylor (1959) [5, p. 314]. Furthermore, the constructive method we use, gives a sharp upper bound for the Hausdorff dimension of the set of exceptional points, for the strong density theorem of Saks.

Introduction

S. Ulam in [3, Problem 146, p. 228] (see also [8, p.78]) posed the following problem: Suppose that E is any Lebesgue measurable set on the real line. Does there exist a real function $\psi : \mathbb{R}^+ \to \mathbb{R}^+$ (depending on E) increasing such that $\lim\limits_{t \to 0^+} \psi(t) = 0$ and

$$\lim_{\substack{x \in I \\ |I| \to 0}} \frac{|I \cap E^c|}{|I| \cdot \psi(|I|)} = 0 \qquad \text{a.e. in } E\,?$$

(here I denotes an interval in $\mathbb{R}$, E^c is the complement of E and $|\cdot|$ stands for the Lebesgue measure).

The affirmative answer to this question was given by S. J. Taylor in [5] and of course this is a strengthening of the Lebesgue density theorem. However, the problem, whether does there exist a function ψ independent of E, has a negative solution (see [5, Theorem 4, p. 312]). Also, in [5], the problem for a similar strengthening of the strong density theorem of Saks was posed (for Saks' strong density theorem, see [4, p. 129]).

Problem ([5, p. 314]). Given an m-dimensional Lebesgue measurable set E, does there exist a real function $\psi(t)$ increasing with $\lim\limits_{t \to 0^+} \psi(t) = 0$ such that

$$\lim_{\substack{x \in R \\ d(R) \to 0}} \frac{|R \cap E^c|}{|R| \cdot \psi(d(R))} = 0 \qquad \text{for a.e. } x \in E\,?$$

(R is an interval in $\mathbb{R}^m$, d is the diameter).

As S. J. Taylor remarked, the direct methods in [5] did not give this strongest form of the density theorem in m-space.

The affirmative answer to this problem was given by S. J. Taylor in [6], assuming the usual form of the strong density theorem and applying an ingenious strengthening of Egoroff's theorem. The answer given is not constructive and gives no information on how the function ψ depends on the set E.

In [9] we give a new constructive answer to the problem of S. J. Taylor in [5], that gives a function ψ not only for a particular set, but for an uncountable class of sets.

To be precise, given a sequence $\{(w_n, u_n)\colon\ n \in \mathbb{N}\}$ in $\mathbb{R}^2$ with $w_n > 0$, $u_n > 0$ for $n \in \mathbb{N}$ such that the series $\sum\limits_{n=1}^{\infty} w_n \cdot u_n$ converges, we say that a set $E \subseteq \mathbb{R}^2$ *belongs to* $\{(w_n, u_n)\colon\ n \in \mathbb{N}\}$ if $E = \bigcup\limits_{n=1}^{\infty} (I_n \times J_n)$, where $I_n \times J_n$ are disjoint intervals in $\mathbb{R}^2$ with $|I_n| = w_n$, $|J_n| = u_n$ for $n \in \mathbb{N}$. Clearly, every bounded open set belongs to some sequence as above and there are uncountably many sets belonging to a given sequence (this kind of definition in one dimension is also considered in [1] for a different purpose). As a result we have that, given a sequence $\{(w_n, u_n)\colon\ n \in \mathbb{N}\}$ as above, we construct a $\psi : \mathbb{R}^+ \to \mathbb{R}^+$ (depending on $\{(w_n, u_n)\colon\ n \in \mathbb{N}\}$) increasing, with $\lim\limits_{t \to 0^+} \psi(t) = 0$ such that for every set E belonging to $\{(w_n, u_n)\colon\ n \in \mathbb{N}\}$

$$\lim_{\substack{x \in (A \times B) \\ d(A \times B) \to 0}} \frac{|(A \times B) \cap E^c|}{|A \times B| \cdot \psi(d(A \times B))} = 0 \qquad \text{for a.e. } x \in E$$

(where $A \times B$ is an interval in $\mathbb{R}^2$). (We remark that the Problem in [5,p.314] is somewhat misstated, since the ratio following the limit is not uniquely determined, for given $d(A \times B)$.)

It should be noted that, this kind of result can be obtained in one dimension by the direct methods of [5]. Also, our method equally works in any dimension, but we restrict to $\mathbb{R}^2$ for notational simplicity.

Furthermore, the constructive method we use, gives a sharp upper bound of the Hausdorff dimension of the set of exceptional points, for the strong density theorem of Saks.

To be precise, in [2] A. S. Besicovitch proved that, given a perfect set $E \subset [0, 1]$ and denoting by φ the sequence $a_1, a_2, \ldots$ of the lengths of interior complementary intervals of E, then the Hausdorff dimension of the set of exceptional points, for the Lebesgue density theorem (for E), is bounded above by the Besicovitch-Taylor index (or exponent of convergence) of the sequence $\{a_n :\ n \in \mathbb{N}\}$ (defined as

$$e_{BT}\{a_n :\ n \in \mathbb{N}\} := \inf\{c > 0 : \sum_{n=1}^{\infty} a_n^c \text{ converges}\} \qquad \text{(see [7, p. 34]))}$$

and this bound is sharp. In this paper we prove that, given a bounded, open set $E \subseteq R$ (R is an open interval in $\mathbb{R}^2$) and if E is written as a countable union of disjoint intervals, i.e. $E = \bigcup\limits_{n=1}^{\infty} (I_n \times J_n)$, then the Hausdorff dimension of the set of exceptional points, for the strong density theorem (for E), is bounded above by

$$1 + \min\{e_{BT}\{|I_n| : \ n \in \mathbb{N}\}, \ \ e_{BT}\{|J_n| : \ n \in \mathbb{N}\}\}$$

and this bound is sharp. Also, working in higher dimensions we have analogous bounds. It should be noted that, clearly this bound is non-trivial, only in case that at least one of the series $\sum_{n=1}^{\infty} |I_n|$, $\sum_{n=1}^{\infty} |J_n|$ converges.

A final remark. The problem, (in the classical proofs of the strong density theorem), of estimating the Hausdorff dimension of the set of exceptional points is that the sections of this set vary. So, there is no effective way of estimating its Hausdorff dimension. The advantage of the present proof is that the set of exceptional points is "placed" as a subset of a cartesian product of $\mathbb{R}$ and a null set. So, after estimating the Hausdorff dimension of this null set, we can estimate that of the set of exceptional points, by standard results of the theory.

References

[1] A.S. Besicovitch, S.J. Taylor, On the complementary intervals of a linear closed set of zero Lebesgue measure, J. Lond. Math. Soc. 29 (1954) 449-459.

[2] A.S. Besicovitch, On density of perfect sets, J. Lond. Math. Soc. 31 (1956) 48-53.

[3] R.D. Mauldin (ed.), The Scottish Book, Birkhauser, Boston, 1981.

[4] S. Saks, Theory of the Integral (trans. L.C. Young), Warsaw, Monografie Matematyczne, 1937.

[5] S.J. Taylor, On strengthening the Lebesgue density theorem, Fund. Math. 46 (1959) 305-315.

[6] S.J. Taylor, An alternative form of Egoroff's theorem, Fund. Math. 48 (1960) 169-174.

[7] C. Tricot, Curves and Fractal Dimension, Springer-Verlag, New York, 1995.

[8] S.M. Ulam, Problems in Modern Mathematics, Dover Phoenix Editions, 2004.

[9] P. Georgopoulos, C. Gryllakis, A sharp upper bound for the Hausdorff dimension of the set of exceptional points for the strong density theorem, J. Math. Anal. Appl. 435, No.2, (2016) 1273-1295.

Panagiotis Georgopoulos: Department of Mathematics, National and Kapodistrian University of Athens, Panepistimioupolis 157-84, Athens, Greece.

E-mail: pgeorgop@math.uoa.gr

Constantinos Gryllakis: Department of Mathematics, National and Kapodistrian University of Athens, Panepistimioupolis 157-84, Athens, Greece.

E-mail: cgryllakis@math.uoa.gr

Harmonic functions, crossed products and approximation properties

Aristides Katavolos

In memory of Dimitris Gatzouras

Abstract

The space of harmonic functions on a locally compact group G is the fixed point space of a certain Markov operator. Its 'quantization', the corresponding fixed point space of operators on L^2G, coincides with the weak* closed bimodule over the group von Neumann algebra generated by this space. We examine the analogous spaces of jointly harmonic functions and their quantized operator bimodules. This leads to two different notions of crossed product of operator spaces by actions of G which coincide when G satisfies a certain approximation property.

The talk is a survey of joint work with M. Anoussis and I. G. Todorov, and of more recent work by D. Andreou.

1 Appetizer: The diagonal problem

Let Γ be a (discrete) group. Any $\phi \in \ell^\infty\Gamma$ defines a multiplication operator $f \mapsto \phi f$: $\ell^2\Gamma \to \ell^2\Gamma$ which we denote by the same symbol ϕ (here $(\phi f)(s) = \phi(s)f(s)$ for $s \in \Gamma$).

This multiplication operator is "diagonal" with respect to the orthonormal basis $\{\delta_s : s \in \Gamma\}$ of $\ell^2\Gamma$.

For $r \in G$, if λ_r is the translation operator $\delta_s \mapsto \delta_{rs}$, the operator $\phi\lambda_r \in \mathcal{B}(\ell^2\Gamma)$ "lives" on the r-th diagonal. Thus the space of operators with finitely many nonzero diagonals is

$$\left\{ \sum_{i=1}^{n} \phi_i \lambda_{r_i} : \phi_i \in \ell^\infty\Gamma, r_i \in \Gamma, n \in \mathbb{N} \right\} .$$

It is not hard to show that this space contains the set of "matrix units" $\{E_{r,s} : r, s \in \Gamma\}$ and hence is dense in $\mathcal{B}(\ell^2\Gamma)$ for the w^*-topology (recall that $\mathcal{B}(\ell^2\Gamma)$ is the dual of the Banach space $\mathcal{S}^1(\ell^2\Gamma)$ of trace class operators). But,

Question 1.1. Is it true that every $X \in \mathcal{B}(\ell^2\Gamma)$ is the "sum of its diagonals"? More precisely, is it true that if $D_r(X) \in \mathcal{B}(\ell^2\Gamma)$ denotes the r-th diagonal of X, then

$$\lim_{F \subset\subset G} \sum_{r \in F} D_r(X) \to X \quad \text{in some sense? } (\subset\subset\text{: finite subset})$$

An incorrect answer appears in more than one classic book...

Question 1.2. What if we introduce "multipliers" $u \in \ell^\infty\Gamma$ (more general than χ_F)? Can we find, for each $X \in \mathcal{B}(\ell^2\Gamma)$, a net (u_i) of finitely supported functions on Γ (possibly depending on X), so that

$$\lim_i \sum_r u_i(r)D_r(X) \to X \text{ ??}$$

2 Harmonic functions

Let $\mu \in M(G)$ be a probability measure on a locally compact group G.

• A function $\phi : G \to \mathbb{C}$ is said to be μ-**harmonic** if

$$\int_G \phi(st)d\mu(t) = \phi(s)\,. \quad \text{Write } \phi \in \mathcal{H}(\mu)\,.$$

We studied the notion of μ-harmonic functions, and its connection to random walks, in a seminar [10] with Dimitris Gatzouras and others.

Here, we will limit ourselves to the functional analysis approach.

Thus a μ-harmonic function ϕ is a fixed point of the map P_μ given by

$$(P_\mu\phi)(s) = \int_G \phi(st)d\mu(t)\,.$$

The map P_μ is positive, unital, w^*-continuous on $L^\infty(G)$.

3 The (classical) Poisson boundary

The space $\mathcal{H}(\mu)$ of μ-harmonic functions is the range of a positive unital projection defined on $L^\infty G$. This projection can be obtained by averaging over iterates of P_μ, as follows:

Note that, since $L^\infty G$ is a dual Banach space, so is $\mathcal{B}(L^\infty G)$, and hence its unit ball is w^*-compact.

Define

$$E_n := \frac{1}{n}\left(I + P_\mu + (P_\mu \circ P_\mu) + \cdots + P_\mu^{n-1}\right) \in \mathrm{ball}\,\mathcal{B}(L^\infty G)$$

and let $E_\mu \in \mathcal{B}(L^\infty G)$ be a w^*-cluster point of $\{E_n\}$.

Then it can be shown that E_μ is a unital positive projection onto $\mathcal{H}(\mu)$.

The space $\mathcal{H}(\mu)$ is not an algebra under pointwise multiplication. But, using the projection E_μ, it can be equipped with an associative multiplication $\diamond$, by defining

$$\phi \diamond \psi := E_\mu(\phi\psi), \quad \phi, \psi \in \mathcal{H}(\mu)\,.$$

Then $(\mathcal{H}(\mu), \diamond)$ becomes a C*-algebra; since it is w^*-closed, it is a von Neumann algebra; and since it is abelian, it is an L^∞ space:

Thus there exists a probability space (Ω, ν), called the **Poisson boundary of** μ so that $\mathcal{H}(\mu) \simeq L^\infty(\Omega, \nu)$.

Let us remark that the von Neumann algebra structure on $(\mathcal{H}(\mu), \diamond)$ is independent of the choice of cluster point for $\{E_n\}$ (see section 5).

4 From Harmonic functions to Harmonic operators

Recall that an $\phi \in L^\infty G$ is a μ-harmonic *function* if

$$\int_G \phi(st)d\mu(t) = \phi(s)\,.$$

If we consider $\phi \in L^\infty G$ as a multiplication operator acting on L^2G, we may write the previous equality as an operator-valued integral

$$P_\mu\phi = \int_G \rho_t\phi\rho_t^{-1}d\mu(t) \in \mathcal{B}(L^2G)$$

which should be interpreted in the "weak" sense. Here ρ is the right regular representation $G \curvearrowright L^2G$ given by

$$(\rho_r f)(s) = \Delta(r)^{1/2}f(sr), \quad f \in L^2(G),\ s, r \in G$$

where $\Delta : G \to \mathbb{R}^+$ is the modular function, defined by $d(tr) = \Delta(r)dt$.

This interpretation allows us to extend the notion of harmonic functions to operators (quantisation):

• Let us call an operator $T \in \mathcal{B}(L^2G)$ **a μ-harmonic operator** if

$$\int_G \rho_t T\rho_t^{-1}d\mu(t) = T\,. \quad \text{Write } \ T \in \widetilde{\mathcal{H}}(\mu)\,.$$

So μ-harmonic operators are *fixed points* of the map

$$\Theta_\mu : \mathcal{B}(L^2G) \to \mathcal{B}(L^2G) : T \to \int_G \rho_t T\rho_t^{-1}d\mu(t)$$

which is an extension of P_μ and a weak* continuous unital and *completely positive* map. (Such maps are sometimes called Markov operators.)

Complete positivity means that, not only does Θ_μ map positive operators to positive operators, but for all n its *n-th ampliation* has the same property: If an $n \times n$ matrix $[T_{ij}]$ of operators defines a positive operator on $(L^2G)^{(n)}$, then $[\Theta_\mu(T_{ij})]$ also defines a positive operator on $(L^2G)^{(n)}$.

5 The non-commutative Poisson boundary

The following construction is due to Arveson [5] and Izumi [7]:

Let $\tilde{E}_\mu \in \mathcal{B}(\mathcal{B}(L^2G))$ be a w^*-cluster point of $\{\tilde{E}_n\}$, where

$$\tilde{E}_n := \frac{1}{n}\left(I + \Theta_\mu + \cdots + \Theta_\mu^{n-1}\right) : \mathcal{B}(L^2G) \to \mathcal{B}(L^2G)\,.$$

(This uses the fact that $\mathcal{B}(L^2G)$ is a dual Banach space, and hence so is $\mathcal{B}(\mathcal{B}(L^2G))$.) Then $\tilde{E}_\mu$ is a unital completely positive projection onto $\widetilde{\mathcal{H}}(\mu)$.

Using $\tilde{E}_\mu$, we can equip $\widetilde{\mathcal{H}}(\mu)$ with an associative multiplication $\diamond$ by defining

$$T \diamond S := \tilde{E}_\mu(TS), \quad T, S \in \widetilde{\mathcal{H}}(\mu)\,.$$

Then the space $\mathcal{N}_\mu := (\widetilde{\mathcal{H}}(\mu), \diamond)$ is a (non abelian) von Neumann algebra and $(\mathcal{H}(\mu), \diamond)$ is an abelian *-subalgebra.

As in the classical case, the von Neumann algebra structure on $\mathcal{N}_\mu$ is independent of the choice of cluster point for $\{\tilde{E}_n\}$: indeed every *completely* positive isometric linear isomorphism between von Neumann algebras must be a *-isomorphism.

The algebra $\mathcal{N}_\mu$ is called the **non-commutative Poisson boundary of** μ.

We would like to find a more "concrete" description. Perhaps the subalgebra $\mathcal{H}(\mu) \simeq L^\infty(\Omega, \nu)$ may provide a "coordinate representation" for $\mathcal{N}_\mu$.

6 Left Ideals of $L^1(G)$ and $\mathrm{VN}(G)$ bimodules

Observe that the preannihilator J_μ of $\mathcal{H}(\mu)$, given by

$$J_\mu := \mathcal{H}(\mu)_\perp = \left\{ f \in L^1 G : \int_G \phi(t) f(t) dt = 0 \, \forall \phi \in \mathcal{H}(\mu) \right\} \subseteq L^1 G$$

is invariant under left translations by G (because $\mathcal{H}(\mu)$ is) so J_μ is a left (convolution) ideal.

More generally, consider any closed left ideal $J \subseteq L^1 G$.

Then $J^\perp \subseteq L^\infty G \subseteq \mathcal{B}(L^2 G)$ is annihilated by the maps Θ_{ν_f} for all $f \in J$ (here $d\nu_f(t) := f(t)dt$); hence $J^\perp$ lies in

$$\begin{aligned} \ker \Theta(J) &= \left\{ T \in \mathcal{B}(L^2(G)) : \int_G \rho_t T \rho_t^{-1} f(t) dt = 0 \ \text{ for all } \ f \in J \right\} \\ &= \bigcap_{f \in J} \ker \Theta_{\nu_f} \,. \end{aligned}$$

But each Θ_{ν_f} commutes with left or right multiplication by left-translation operators $\{\lambda_t, \ t \in G\}$ on $L^2(G)$. Thus $\ker \Theta(J)$ is a *bimodule* over the w^*-closed linear span of $\{\lambda_t, \ t \in G\}$, which is known as the *von Neumann algebra* $\mathrm{VN}(G)$ *of* G. It follows that $\ker \Theta(J)$ also contains the w^*-closed space

$$\mathrm{Bim}(J^\perp) := \overline{\mathrm{span}}^{w^*} \{ \phi \lambda_t : \phi \in J^\perp, t \in G \}$$

which is a bimodule over $\mathrm{VN}(G)$ (since $\lambda_s \phi \lambda_t = \phi_s \lambda_{st}$ where $\phi_s(t) = \phi(s^{-1}t)$).

Thus for every closed left ideal $J \subseteq L^1 G$ we have the inclusion

$$\mathrm{Bim}(J^\perp) \subseteq \ker \Theta(J) \,.$$

We think of the elements of $\mathrm{Bim}(J^\perp)$ as w^*-limits of "polynomials" $\sum \phi_i \lambda_{t_i}$ whose coefficients $\phi_i \in L^\infty G$ are annihilated by all $f \in J$. On the other hand, $\ker \Theta(J)$ consists of all operators annihilated by $\{\Theta_{\nu_f}, f \in J\}$.

When can we approximate such operators by suitable polynomials $\sum \phi_i \lambda_{t_i}$?

Theorem 6.1. *If G has the Approximation Property AP of Haagerup-Kraus, then the equality*

$$\mathrm{Bim}(J^\perp) = \ker \Theta(J) \,.$$

holds for every left ideal $J \subseteq L^1(G)$.

(See section 8 for the Approximation Property).

Thie validity of this equality was first proved for G abelian, or compact, or weakly amenable discrete with M. Anoussis and I.G. Todorov [3]. The general case was then proved by J. Crann and M. Neufang [4].

7 Application to jointly harmonic operators

Given a family $\Lambda \subseteq M(G)$ of (complex valued) measures, the set $\mathcal{H}(\Lambda)$ of *jointly harmonic functions* is the set $\bigcap_{\mu\in\Lambda} \mathcal{H}(\mu)$ consisting of all $\phi \in L^\infty G$ which are μ-harmonic for all $\mu \in \Lambda$.

Correspondingly, we define the set of all *jointly harmonic operators* to be

$$\begin{aligned}\widetilde{\mathcal{H}}(\Lambda) &:= \{T \in \mathcal{B}(L^2(G)) : \mu\text{-harmonic for all } \ \mu \in \Lambda\} \\ &= \{T \in \mathcal{B}(L^2(G)) : \Theta(\mu)(T) = T \text{ for all } \ \mu \in \Lambda\}\,.\end{aligned}$$

Clearly, $\widetilde{\mathcal{H}}(\Lambda) \supseteq \mathrm{Bim}(\mathcal{H}(\Lambda))$ where $\mathrm{Bim}(\mathcal{H}(\Lambda))$ is the w^*-closed linear space generated by $\{\phi\lambda_t : \phi \in \mathcal{H}(\Lambda), t \in G\}$.

Theorem 6.1 is applicable not only to ideals $J \subseteq L^1(G)$ which are preannihilators of μ-harmonic functions, but also to preannihilators of jointly Λ-harmonic functions:

Theorem 7.1. *Suppose G has the Approximation Property.*
For any $\Lambda \subseteq M(G)$,

$$\widetilde{\mathcal{H}}(\Lambda) = \mathrm{Bim}(\mathcal{H}(\Lambda))\,.$$

Remark In the special case of functions which are μ harmonic for a probability measure, the equality $\widetilde{\mathcal{H}}(\mu) = \mathrm{Bim}(\mathcal{H}(\mu))$ holds for all groups. This was shown for discrete groups by M. Izumi [7], and then for general locally compact groups by W. Jaworski and M. Neufang [8] using completely different methods.

The crucial point, in this special case, is that the space $\mathcal{H}(\mu)$ is linearly and covariantly completely isometrically isomorphic to a von Neumann algebra, namely $L^\infty(\Omega, \nu)$ where (Ω, ν) is the Poisson boundary.

8 Interlude: the approximation property AP

Very roughly, a locally compact G has the approximation property AP of Haagerup-Kraus when the Fourier algebra $A(G)$ contains an (unbounded) approximate identity of a weak form. The *Fourier algebra of G* consists of all functions $u : G \to \mathbb{C}$ of the form $u(s) = \langle \lambda_s f, g\rangle$ where $f, g \in L^2G$. Every such fuction defines a bounded mutiplier $M_u : \mathrm{VN}(G) \to \mathrm{VN}(G)$ satisfying $M_u(\lambda_s) = u(s)\lambda_s$ for all $s \in G$.

The following can be taken as the definition: G has the AP if and only if there is a net (u_i) of compactly supported functions in $A(G)$ such that (M_{u_i}) converges in the stable point-weak* topology to the identity, i.e. $(M_{u_i} \otimes \mathrm{id})(a) \to a$ weak* for all $a \in \mathrm{VN}(G)\bar{\otimes}\mathcal{B}(\ell^2)$ [6, Theorem 1.9].

The AP is a weak form of amenability.

Examples *Groups with the AP: Amenable groups, such as abelian or compact groups, but also some non-amenable, such as* $\mathbb{F}_n$.
Groups without the AP: $SL(3,\mathbb{Z})$, $SL(3,\mathbb{R})$.

Under the AP, we can answer Question 1.2:

Proposition 8.1. *If* Γ *is a discrete group with the AP, every operator in* $\mathcal{B}(\ell^2\Gamma)$ *can be* w^**-approximated by linear combinations of its own diagonals.*

Indeed the M_{u_i} mentionned above extend to operators defined on the whole of $\mathcal{B}(L^2G)$ and provide the required multipliers.

9 Change of perspective: The crossed product

Let us return to the space $\mathcal{H}(\Lambda)$ of functions in $L^\infty G$ which are jointly harmonic for a family Λ of complex measures on G, together with its "quantised" cousin $\widetilde{\mathcal{H}}(\Lambda)$ of jointly harmonic operators. Note that G acts on $\mathcal{H}(\Lambda)$ by left translations. We wish to use $\mathcal{H}(\Lambda)$ together with this action to describe the space $\widetilde{\mathcal{H}}(\Lambda)$.

More generally: Let $\mathcal{V} \subseteq \mathcal{B}(H)$ be a w^*-closed linear space of operators on some Hilbert space H (a *dual operator space*) and let $s \mapsto \alpha_s$ be an action of G on $\mathcal{V}$ by weak* continuous complete isometries.

We wish to represent both G and $\mathcal{V}$ simultaneously and covariantly on the same space. For this, we "create more space" by enlarging H to accomodate both:

Consider

$$\mathcal{V}\bar{\otimes}L^\infty G \subseteq \mathcal{V}\bar{\otimes}\mathcal{B}(L^2G) \subseteq \mathcal{B}(H \otimes L^2G)$$

(we use $\bar{\otimes}$ for the w^*-closure of the algebraic tensor product).

We represent $\mathcal{V}$ on $H\otimes L^2G$ as follows: thinking of $\mathcal{V}\bar{\otimes}L^\infty G$ as consisting of $\mathcal{V}$-valued L^∞ functions on G, we associate to each $v \in \mathcal{V}$ the function $s \mapsto \alpha_s^{-1}(v)$.

More precisely, for each $v \in \mathcal{V}$, we define $\pi_\alpha(v) \in \mathcal{V}\bar{\otimes}L^\infty(G)$ by duality:

$$\langle \pi_\alpha(v), \omega \otimes h\rangle := \int_G \langle \alpha_s^{-1}(v), \omega\rangle h(s)ds, \qquad \omega \in \mathcal{V}_*, h \in L^1(G)\,.$$

(Here $\mathcal{V}_*$ is the space of all w^*-continuous linear forms on $\mathcal{V}$, and we are using the fact that the projective tensor product of simple tensors of $\mathcal{V}_*$ and $L^1(G)$ has $\mathcal{V}\bar{\otimes}L^\infty G$ as its dual.)

We also define a map

$$\tilde{\lambda} : G \to \mathcal{B}(H \otimes L^2G) : s \mapsto \tilde{\lambda}_s := \mathrm{Id}_H \otimes \lambda_s\,.$$

So we have the representations

$$\begin{aligned}\pi_\alpha &: \mathcal{V} \to \mathcal{V}\bar{\otimes}\mathcal{B}(L^2G) \subseteq \mathcal{B}(H \otimes L^2G)\\ \tilde{\lambda} &: G \to \mathcal{B}(H \otimes L^2G).\end{aligned}$$

The point is that now the action α becomes "inner": it is implemented by the unitary group $\tilde{\lambda}$:

$$\pi_\alpha(\alpha_s(v)) = \tilde{\lambda}_s\pi_\alpha(v)\tilde{\lambda}_s^{-1}\,.$$

This setup allows us to define *two versions* of the crossed product:

• The **spatial crossed product** $\mathcal{V} \rtimes_\alpha G$ is defined to be the weak* closed subspace of $\mathcal{V}\bar{\otimes}\mathcal{B}(L^2G)$ generated by all "polynomials" in $\{\tilde{\lambda}_s : s \in G\}$ with "coefficients" from $\pi_\alpha(\mathcal{V})$: it is the weak* closed space

$$\mathcal{V} \rtimes_\alpha G := \overline{\operatorname{span}\{\pi_\alpha(v)\tilde{\lambda}_s,\ v \in \mathcal{V}, s \in G\}}^{w*} \subseteq \mathcal{V}\bar{\otimes}\mathcal{B}(L^2G).$$

• The **Fubini crossed product** $\mathcal{V} \rtimes_\alpha^F G$ is defined to be the following fixed point subspace of $\mathcal{V}\bar{\otimes}\mathcal{B}(L^2G)$:

$$\mathcal{V} \rtimes_\alpha^F G := \{T \in \mathcal{V}\bar{\otimes}\mathcal{B}(L^2G) : (\alpha_s \otimes \mathrm{Ad}\rho_s)(T) = T\ \forall s \in G\}$$

Here $\tilde{\alpha}_s := \alpha_s \otimes \mathrm{Ad}\rho_s$ acts on a simple tensor $T = x \otimes y$ as follows:
$\tilde{\alpha}_s(T) = \alpha_s(x) \otimes \rho_s y \rho_s^{-1}$.

It is not hard to see that $\pi_\alpha(\mathcal{V})$ and $\tilde{\lambda}(G)$ are both elementwise fixed by the action $\tilde{\alpha}$; hence so is the spatial crossed product generated by them;

$$\mathcal{V} \rtimes_\alpha G \subseteq \mathcal{V} \rtimes_\alpha^F G\,.$$

But do we have equality? In other words, can every $\tilde{\alpha}$-fixed point be w^*-approximated by "polynomials" of the above form?

It is a classical result (see, for example, [9, Corollary X.1.22]) that these two crossed products coincide in case $\mathcal{V}$ is *a von Neumann algebra*. However, for more general dual operator spaces, they can be distinct.

Theorem 9.1 (D. Andreou, [1]). *The equality $\mathcal{V} \rtimes_\alpha G = \mathcal{V} \rtimes_\alpha^F G$ holds for all dual operator spaces $\mathcal{V}$ if and only if the group G has the AP.*

Note that the "if" direction was also proved by Crann - Neufang [4] using a different approach.

This Theorem can be viewed as a dynamical characterization of the AP.

10 Bimodules and Crossed products

We now apply these concepts to the Kernel-Bimodule problem. The key is the following:

In the special case where $\mathcal{V} = L^\infty G$ and G acts by left translation (we write $G \overset{\alpha_G}{\curvearrowright} L^\infty G$) both crossed products can be represented on $L^2(G)$:

Proposition 10.1 (D. Andreou, [1]). *There is an isometric normal *-morphism $\Psi : \mathcal{B}(L^2G) \to \mathcal{B}(L^2G)\bar{\otimes}\mathcal{B}(L^2G)$ such that: for any closed left ideal J of $L^1(G)$, we have*

$$\mathrm{Bim}(J^\perp) \overset{\Psi}{\simeq} J^\perp \rtimes_{\alpha_G} G \quad \textit{and} \quad \ker\Theta(J) \overset{\Psi}{\simeq} J^\perp \rtimes_{\alpha_G}^F G.$$

Therefore, applying Theorem 9.1, we obtain a conceptually different proof of Theorem 6.1:

Under the AP, the equality $\mathrm{Bim}(J^\perp) = \ker\Theta(J)$ holds for all closed left ideals $J \subseteq L^1G$. In particular, for any $\Lambda \subseteq M(G)$, the space $\widetilde{\mathcal{H}}(\Lambda)$ of jointly harmonic operators is (isomorphic *to) the spatial crossed product of the space $\mathcal{H}(\Lambda)$ of jointly harmonic functions by the translation action of G.*

Concluding Remarks We have seen that **if** a group G has the AP, **then**

$$J^{\perp} \rtimes_{\alpha_G} G = J^{\perp} \rtimes^{F}_{\alpha_G} G \quad \text{for all closed left ideals } J \text{ of } L^1(G) \tag{$*$}$$

but we do not know whether the converse holds:

Question: Is the AP necessary for the validity of ($*$)?

Or is some weaker approximation property sufficient?

Or is ($*$) valid for all locally compact bgroups G?

Acknowledgment. Thanks are due to D. Andreou and M. Anoussis for their help in the preparation of this survey.

References

[1] D. Andreou, *Crossed products of dual operator spaces and a characterization of groups with the AP*, J. Op. Theory (to appear).

[2] D. Andreou, *Crossed products of dual operator spaces by locally compact groups*, Studia Mathematica **258** (2021), 241-267.

[3] M. Anoussis, A. Katavolos, I. G. Todorov, *Bimodules over $VN(G)$, harmonic operators and the non-commutative Poisson boundary*, Studia Mathematica **249** (2019), 193-213.

[4] J. Crann, M. Neufang, *A non-commutative Fejér theorem for crossed products, the approximation property, and applications*, arXiv:1901.08700. International Mathematics Research Notices. (2020).

[5] W. Arveson, *Noncommutative Poisson boundaries*, lecture notes (2004).

[6] U. Haagerup, J. Kraus, *Approximation properties for group C*-algebras and group von Neumann algebras*, Trans. Amer. Math. Soc., **344** (1994), no. 2, 667-699.

[7] M. Izumi, *Non-commutative poisson boundaries*. In M. Kotani, T. Shirai, and T. Sunada, editors, *Discrete Geometric Analysis*, volume 347 of Contemporary Mathematics. Amer. Math. Soc. (2004).

[8] W. Jaworski, M. Neufang, *The Choquet-Deny equation in a Banach space* Canadian J. Math. 59 (2007), 795-827.

[9] M. Takesaki, *Theory of operator algebras. II*, volume 125 of *Encyclopaedia of Mathematical Sciences*. Springer-Verlag, Berlin (2003).

[10] M. Anoussis, D. Cheliotis, D. Gatzouras, A. Katavolos (organizers), *Random walks on groups*, NKUA seminar (in Greek, 2017).

Aristides Katavolos: Department of Mathematics, National and Kapodistrian University of Athens, Panepistimioupolis 157-84, Athens, Greece.

E-mail: akatavol@math.uoa.gr

Simultaneous tiling

Mihail N. Kolountzakis[*]

Dedicated to the memory of Dimitris Gatzouras

Abstract

We discuss problems of simultaneous tiling. This means that we have an object (set, function) which tiles space with two or more different sets of translations. The most famous problem of this type is the Steinhaus problem which asks for a set simultaneously tiling the plane with all rotates of the integer lattice as translation sets.

Contents

1 Tiling by translation

For the purposes of this paper [1] tiling is by translation only [16]. We have an object T (the *tile*) which may be a set or a function on some abelian group G (usually the Euclidean space but it may be $\mathbb{Z}^d$ or a finite group) which we are translating around

[*]Supported by the Hellenic Foundation for Research and Innovation, Project HFRI-FM17-1733 and by grant No 4725 of the University of Crete.

[1]This is a survey paper which contains no new results. Many of the results and questions in this paper are from [21] on parts of which this paper is heavily based.

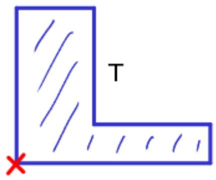


Figure 1: An L-shaped tile. The red point is the origin

by a set of translations Λ, in such a way that everything in the group G is covered exactly once, with the possible exception of a set of zero Haar measure, to account for such irrelevant things such as boundaries overlapping, which we generally do not care about. One convenient way to define tiling by a function f (which can be an indicator

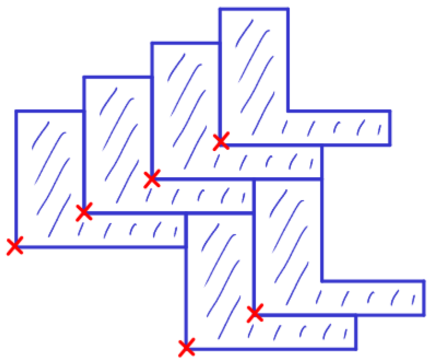


Figure 2: A translational tiling by the L-shaped region. The red points are the translation set.

function, if we want tiling by a set) when translated at the locations Λ is to demand that

$$\sum_{\lambda\in\Lambda} f(x-\lambda) = \text{const.}, \tag{1.1}$$

for almost all $x \in G$. To avoid most issues of convergence it makes sense to ask that $f \geq 0$, though some interesting problems do arise with signed f [20].

2 Tiling in Fourier space

It is easy to see that (1.1) may be rewritten as a convolution

$$f * \delta_\Lambda = \text{const.} \tag{2.1}$$

where $\delta_\Lambda = \sum_{\lambda\in\Lambda} \delta_\lambda$ is the measure that encodes the locations Λ by placing a unit mass on each of them. Taking the Fourier Transform of this we obtain

$$\widehat{f\delta_\Lambda} = C\delta_0. \tag{2.2}$$

This implies that the tempered distribution $\widehat{\delta_\Lambda}$ is supported on the zeros of $\widehat{f}$ plus the origin

$$\operatorname{supp}\widehat{\delta_\Lambda} \subseteq \left\{\widehat{f}=0\right\} \cup \{0\}. \tag{2.3}$$

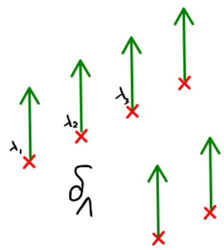


Figure 3: The collection δ_Λ of point masses that encodes the set Λ

Let us now restrict ourselves to the case of $G = \mathbb{R}^d$ and $\Lambda \subseteq \mathbb{R}^d$ being a lattice $\Lambda = A\mathbb{Z}^d$, where A is a non-singular $d \times d$ matrix. The Poisson summation formula reads

$$\widehat{\delta_\Lambda} = \frac{1}{|\det A|}\delta_{\Lambda^*}$$

in this case, where $\Lambda^* = A^{-\top}\mathbb{Z}^d$ is the dual lattice of Λ, so the tiling of f with Λ becomes equivalent to

$$\widehat{f}(\lambda^*) = 0 \quad \text{for all } \lambda^* \in \Lambda^* \setminus \{0\}.$$

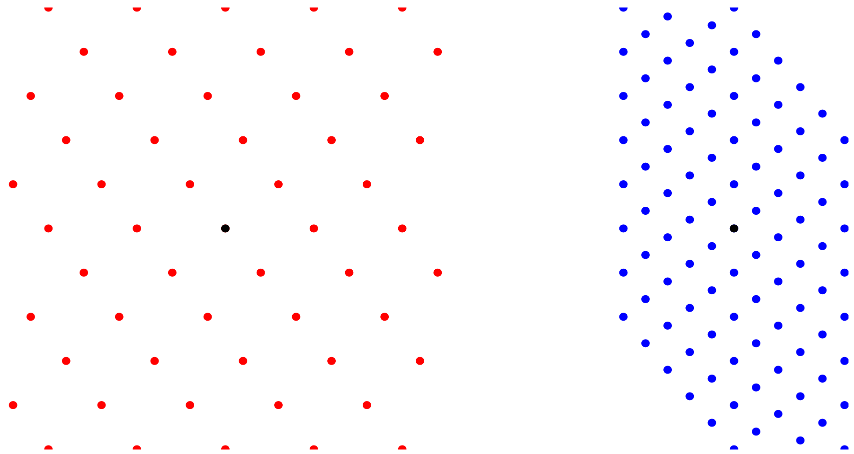


Figure 4: A lattice Λ and its dual Λ^*

3 The Steinhaus tiling problem

In the *Steinhaus tiling problem* we are seeking a tile that can tile *simultaneously* with many different sets of translations. The most important case is: can we find a subset of the plane which can tile (by translations) with all rotates of the integer lattice $\mathbb{Z}^2$? In some sense we are asking for a set in the plane that can behave simultaneously like all these rotated squares (Fig. 5). There are two major variations of the Steinhaus problem: the *measurable* and the *set-theoretic* case. In the measurable case we demand our tile to be a Lebesgue measurable subset of $\mathbb{R}^d$ and we are, at the same time, relaxing our requirements and are allowing a subset of measure 0 of space not to be covered exactly once by the translates of the tile. In the set-theoretic case we allow the tile to

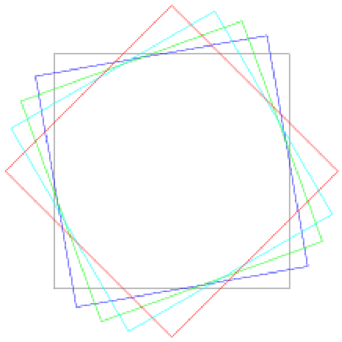


Figure 5: The rotated squares are fundamental domains of all rotates of $\mathbb{Z}^2$

by any subset and we typically ask that every point is covered exactly once, allowing no exceptions.

Komjáth [22] answered the Steinhaus question in the affirmative in $\mathbb{R}^2$ when tiling by all rotates of the set $B = \mathbb{Z} \times \{0\}$ showing that there are such Steinhaus sets (but such a set A cannot be measurable as was shown recently in [18]). Sierpinski [26] showed that a bounded set A which is either closed or open cannot have the lattice Steinhaus property (that is, intersect all rigid motions of $\mathbb{Z}^2$ at exactly one point – another way to say that A tiles precisely with all rotates of $\mathbb{Z}^2$). Croft [5] and Beck [1] showed that no bounded and measurable set A can have the lattice Steinhaus property (but see also [23]). Kolountzakis [14, 13] and Kolountzakis and Wolff [19] proved that any measurable set in the plane that has the measurable Steinhaus property must necessarily have very slow decay at infinity (any such set must have measure 1). In [19] it was also shown that there can be no measurable Steinhaus sets in dimension $d \geq 3$ (tiling with all rotates $\rho\mathbb{Z}^d$, where ρ is in the full orthogonal group) a fact that was also shown later by Kolountzakis and Papadimitrakis [17] by a very different method. See also [3, 24, 4, 27]. Kolountzakis [15] looks at the case where we are only asking for our set to tile with *finitely many* lattices, not all rotates as in the original problem, which we are also doing in this paper. In a major result Jackson and Mauldin [11, 10] proved the existence of Steinhaus sets in the plane which tile with all rotates of $\mathbb{Z}^2$ (not necessarily measurable). Their method does not extend to higher dimension $d \geq 3$. See also [25, 12]. It was also shown in [18] that a set A which tiles with all rotates of a finite set B cannot be measurable.

4 Steinhaus problem in Fourier space

Most of the results on the measurable Steinhaus problem start by observing that if $E \subseteq \mathbb{R}^2$ is Steinhaus then every rotate $R_\theta E$ of E tiles with $\mathbb{Z}^2$, which means that for every angle θ the Fourier Transform $\widehat{\mathbf{1}_{R_\theta E}}$ vanishes on $\mathbb{Z}^2 \setminus \{0\}$ since $\mathbb{Z}^2$ is the dual lattice of itself. This implies that $\widehat{\mathbf{1}_E}$ vanishes on all rotates of $\mathbb{Z}^2$. In other words $\widehat{\mathbf{1}_E}$ vanishes on all circles centered at the origin that go through at least one integer lattice point. The number of these circles is large. There are a little less than $O(R^2)$ such

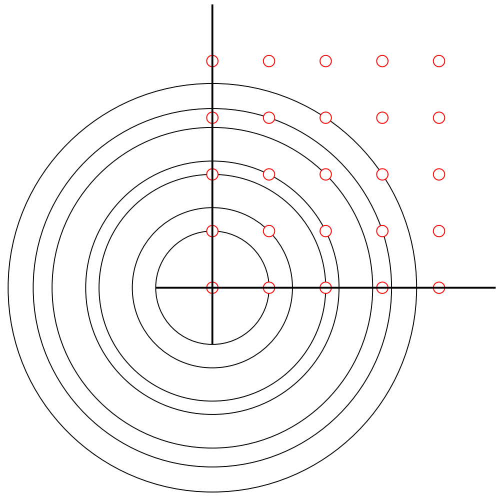


Figure 6: The Fourier Transform of any Steinhaus set must vanish on all circles centered at the origin that go through at least one integer lattice point

circles of radius $\le R$. Many zeros of a function sometimes imply decay at infinity, and, by the usual uncertaintly principle (both f and $\widehat{f}$ cannot decay fast at infinity), since $\widehat{\mathbf{1}_E}$ is small at infinity it follows that $\mathbf{1}_E$ is large (e.g., E cannot be bounded).

5 Allowing functions instead of sets in the Steinhaus problem

Let us now relax our requirements and allow our tile to be a function instead of a set (instead of indicator function, in other words). Satisfying the requirements of the Steinhaus tiling problem with a function is generally much easier than with a set. The problem becomes interesting only if one asks for further properties that this function should have. Therefore we try to find a function with *small support*, or to prove that the support of such a function must necessarily be large. Asking for f to have a small support goes against f having the ability to tile space, especially with many different sets of translations T. The reason is that for f to tile by translations with T its Fourier transform must contain a rich set of zeros [16]. This set of zeros must be able to support the Fourier transform of the measure $\delta_T = \sum_{t\in T} \delta_t$ (which encodes the set of translations). By the well known uncertainty principle in harmonic analysis a rich set of zeros for $\widehat{f}$ usually requires (in various different senses) a large support for f [9].

It is very easy to take $\widehat{f}$ to vanish on the required circles, but one must do it in a way that ensures that f is itself *small* in some sense, such as the diameter of its support or the volume of its support.

6 Small diameter of the support: lower bounds

The first thing that comes to mind is to take f to be a convolution. It takes a moment to verify that if f tiles with a set of translates T then so does $g * f$ for any $g \in L^1(\mathbb{R}^d)$. One can either verify this by checking the definition of tiling for $g * f$ or observe that tiling is a condition that can be checked on the Fourier side [16] and $\widehat{g * f} = \widehat{g} \cdot \widehat{f}$ has an even richer set of zeros that $\widehat{f}$.

So, since $\widehat{f}$ has to vanish on the dual lattices $\Lambda_i^* \setminus \{0\}$ we can take

$$f = \mathbf{1}_{D_1} * \mathbf{1}_{D_2} * \cdots * \mathbf{1}_{D_N}, \tag{6.1}$$

where D_i is a fundamental parallelepiped of Λ_i. Since $D_i + \Lambda_i$ is a tiling it follows that $\widehat{\mathbf{1}_{D_i}}$ vanishes on $\Lambda_i^* \setminus \{0\}$ and that f vanishes on their union and hence tiles with all Λ_i. This can be slightly generalized by taking, instead if the indicator functions $\mathbf{1}_{D_i}$ any function f_i that tiles with Λ_i

$$f = f_1 * f_2 * \cdots f_N. \tag{6.2}$$

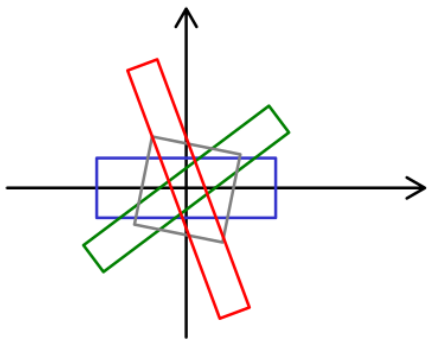


Figure 7: The fundamental domains of several lattices. A constant fraction of them project to a set of large diameter onto one of the coordinate axes.

The following observation (see detailed proof in [21]) was already made in [19] in the case $f_i = \mathbf{1}_{D_i}$.

Theorem 1. *If $\Lambda_1, \ldots, \Lambda_N$ are lattices in $\mathbb{R}^d$ of volume $c_1 \le \operatorname{vol} \Lambda_i$ and $f = f_1 * f_2 * \cdots * f_N$ then*

$$\operatorname{diam} \operatorname{supp} f \ge C_d N. \tag{6.3}$$

The reason is that a constant fraction of the supports of the f_i project onto a constant fraction of their diameter onto *some* line, say one of the axes. This implies (obvious if the f_i are nonnegative; one needs the Titchmarsh convolution theorem in the general case) that so does the support of the convolution $f = f_1 * \cdots * f_N$ (shown in Fig. 7 for the f_i being the indicator functions of fundamental parallelepipeds of the lattices).

If the lattices Λ_i satisfy some "roundness" assumption, e.g. if each Λ_i is assumed to have a fundamental domain of diameter bounded independent of N (as in the important case when all the lattices are rotates of $\mathbb{Z}^d$), then the convolution tile (6.1) has diameter which is also at most $C \cdot N$.

On the other hand we have the following rather general lower bound for the diameter of the support [19] assuming only a certain genericity assumption (6.4) on the Λ_i.

Theorem 2. *If $\Lambda_1, \ldots, \Lambda_N \subseteq \mathbb{R}^d$, $d \geq 1$, are lattices of volume equal to 1 such that*

$$\Lambda_i \cap \Lambda_j = \{0\} \quad \textit{for all } i \neq j, \tag{6.4}$$

then if f tiles with all these lattices we have

$$\operatorname{diam} \operatorname{supp} f \geq C_d N^{1/d}. \tag{6.5}$$

The main question is therefore:

Question 1. *Can the gap between the lower bound* (6.5) *and the linear upper bound $O(N)$ achievable by the convolution tile* (6.1) *(in the case of "round" lattices, having fundamental domains bounded in diameter by a constant) be bridged?*

Are there examples of lattices Λ_i, $i = 1, 2, \ldots, N$, satisfying (6.4) *and a non-zero function $f \in L^1(\mathbb{R}^d)$ that tiles with all Λ_i and such that*

$$\operatorname{diam} \operatorname{supp} f = o(N)?$$

In other words, do there exist collections of lattices for which a common tile f can be found which is diameter-wise more efficient than the convolution construction (6.1)*?*

7 A case of large diameter

We observe now [21] that for some collections of lattices the linear upper bound cannot be improved. The lattices given are both "round" (have a fundamental domain bounded independent of N) and satisfy the genericity assumption (6.4). There are however *collinearities* so, in some sense, this is not a generic situtation.

Theorem 3. *For $d \geq 1$ and for each N there are lattices $\Lambda_1, \ldots, \Lambda_N \subseteq \mathbb{R}^d$, of volume 1, such that if $f \in L^1(\mathbb{R}^d)$, $\int f \neq 0$, tiles with all of them then*

$$\operatorname{diam} \operatorname{supp} f \geq C_d N.$$

Proof. We give the proof in the case $d = 2$. It works with obvious changes in all dimensions $d > 2$ and it is even easier in dimension $d = 1$.

Take Λ_i^* to be generated by the two vectors

$$u_i = (0, a_i), v_i = (1/a_i, 0),$$

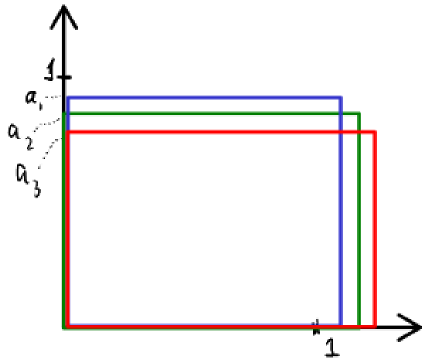


Figure 8: The fundamental rectangles of the lattices of Theorem 7 which have only very long common tiles.

where the numbers $a_1, \ldots, a_N$ are linearly independent over $\mathbb{Q}$ and

$$0.9 < a_i < 1.$$

If f tiles with all Λ_i then $\widehat{f}$ vanishes on all points of the form

$$(0, k \cdot a_i), \quad i = 1, 2, \ldots, N, \quad k \in \mathbb{Z} \setminus \{0\}.$$

Since all these points are different it follows that the density of zeros on the y-axis is $\geq C \cdot N$. This implies that

$$\operatorname{diam} \operatorname{supp} \pi_2(f) \geq C \cdot N$$

(say, by Jensen's formula) where $\pi_2(f)$ is the one-variable function

$$\pi_2(f)(y) = \int_{\mathbb{R}} f(x, y)\, dx.$$

(This is not an identically zero function by our assumption on the integral of f.) This in turn implies

$$\operatorname{diam} \operatorname{supp} f \geq C \cdot N.$$

$\square$

8 Small volume of the support

Another measure of the size of the support is its volume. Can we construct a common tile f for the lattices Λ_i such that $|\operatorname{supp} f|$ is small?

In the case of f given by (6.1) it is clear that

$$\operatorname{supp} f = D_1 + D_2 + \ldots + D_N.$$

To keep things concrete let us assume that all $|D_i| = 1$ in (6.1) (unimodular lattices). Then the Brunn-Minkowski inequality [7] says that

$$|\operatorname{supp} f| = |D_1 + \cdots + D_N| \geq \left(|D_1|^{1/d} + \cdots + |D_N|^{1/d}\right)^d \geq N^d.$$

This lower bound

$$|\mathrm{supp}\, f| \geq CN^d$$

clearly holds also for functions of the form

$$f = f_1 * f_2 * \cdots * f_N, \quad f_i \geq 0, \tag{8.1}$$

where for all $i = 1, 2, \ldots, N$ we assume that the *nonnegative* function f tiles with Λ_i.

We have proved [21]:

Theorem 4. *For any collection of lattices $\Lambda_1, \ldots, \Lambda_N$ in $\mathbb{R}^d$ of volume at least 1 and any common tile f for them of the form*

$$f = f_1 * f_2 * \cdots * f_N, \quad f_i \geq 0,$$

with f_i tiling with Λ_i, we have

$$|\mathrm{supp}\, f| \geq N^d.$$

But when the functions f are signed (or complex) we only have

$$\mathrm{supp}\, f \subseteq \mathrm{supp}\, f_1 + \cdots \mathrm{supp}\, f_N,$$

not necessarily equality, which brings us to the next question.

Question 2. *If f is given by* (8.1)*, is it true that*

$$|\mathrm{supp}\, f| \geq CN^d? \tag{8.2}$$

If one requires that the lattices $\Lambda_1, \Lambda_2, ..., \Lambda_N \subset \mathbb{R}^d$ have the same volume, say 1, and the sum $\Lambda_1^* + \Lambda_2^* + ... + \Lambda_N^*$ of their dual lattices is direct, then, by [15, Theorem 2], they possess a measurable common almost fundamental domain E (generally unbounded).See Fig. 9.

In this case, $|E| = \mathrm{vol}(\Lambda_i) = 1$. So then one can take $f = \mathbf{1}_E$, which tiles with all Λ_i, $i = 1, 2, ..., N$, with $|\mathrm{supp}\, f| = |E| = 1$.

Motivated by the previous observation, but now dropping the equal volume assumption, we ask the following:

Question 3. *Consider the lattices $\Lambda_1, \Lambda_2, ..., \Lambda_N$, with $\frac{1}{2} \leq \mathrm{vol}(\Lambda_i) \leq 2$. Is there a function f that tiles with all Λ_i, such that*

$$|\mathrm{supp}\, f| = o(N^d)?$$

Question 4. *In the case when $\Lambda_1, \ldots, \Lambda_N$ all have volume 1 and satisfy some sort of genericity condition, such as $\Lambda_1^* + \cdots \Lambda_N^*$ being a direct sum, as in [15, Theorem 2], can the common fundamental domain of the Λ_i be bounded? In the construction of [15, Theorem 2] the unboundedness is unavoidable, but is it in the nature of things?*

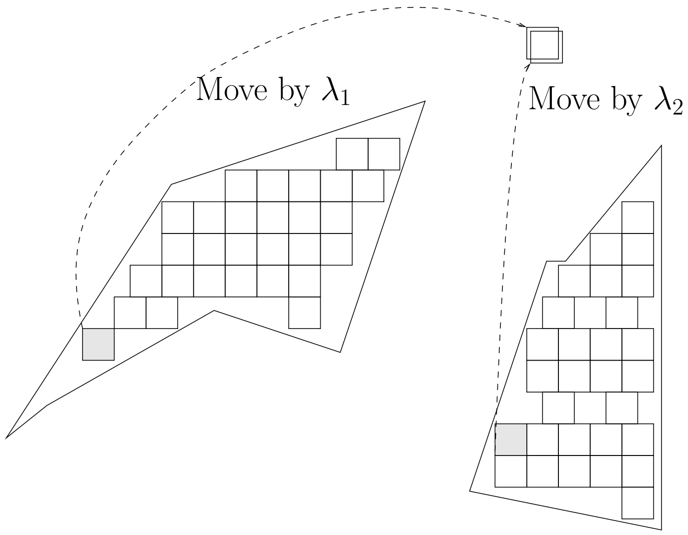


Figure 9: How to rearrange the fundamental domains of two lattices so that they agree almost everywhere [15, Theorem 2]. One breaks up the two domains into smaller and smaller parts, then moves each by a vector in its own lattice so that they agree almost completely.

9 Small length of the support in $d = 1$

In the simplest case in dimension $d = 1$, and for two lattices only, a basic question is to ask if the convolution (8.1) is best in terms of the length of the support. Here we can give [21] a simple lower bound assuming a nonnegative function.

Theorem 5. *Suppose the nonnegative $f : \mathbb{R} \to \mathbb{R}^{\ge 0}$ is measurable and tiles with both $\Lambda_1 = \mathbb{Z}$ and with $\Lambda_2 = \alpha\mathbb{Z}$, where $\alpha \in (0, 1)$:*

$$\sum_{n\in\mathbb{Z}} f(x - n) = 1, \quad \sum_{n\in\mathbb{Z}} f(x - n\alpha) = \frac{1}{\alpha}, \quad \text{for almost every } x \in \mathbb{R}. \tag{9.1}$$

Then

$$|\mathrm{supp}\, f| \ge \left\lceil \frac{1}{\alpha} \right\rceil \alpha \ge 2\alpha. \tag{9.2}$$

Remark 2. If we assume the first equation in (9.1) then the constant in the second equation is forced to be $1/\alpha$. This is because $\int f = 1$ (from the first equation), so repeating f at a set of translates of density $1/\alpha$ will give a constant (assuming it tiles) at that level.

Remark 3. Notice that if α is just a little less than 1 then (9.2) gives a lower bound of 2α, which shows that the convolution $\mathbf{1}_{[0,1]} * \mathbf{1}_{[0,\alpha]}$ is almost optimal in this case, having support of size $1 + \alpha$.

But if, on the other hand, α is just over $1/2$ then the lower bound is just over 1 but the convolution upper bound is just over $3/2$, a considerable gap.

Proof. From the first equation in (9.1) it follows that $f(x) \le 1$ for almost every x. For the second equation to be true it therefore follows that for almost every $x \in \mathbb{R}$ there are at least $\lceil 1/\alpha \rceil$ different values of $n \in \mathbb{Z}$ such that $f(x - n\alpha) > 0$. Using this for almost all $x \in [0, \alpha)$ (which ensures that for different x the locations $x - n\alpha$ are also different) gives (9.2). $\square$

Question 5. *What is the least possible length of the support of f for a nonnegative f that tiles with both $\mathbb{Z}$ and $\alpha\mathbb{Z}$?*

10 Very small diameter of the support with relations among the lattices

If we have N lattices

$$\Lambda_1, \ldots, \Lambda_N \subseteq \mathbb{R}^d$$

we can find a function that tiles with them all, namely the function f in (6.1). If our lattices are assumed to each have a fundamental domain bounded by ~ 1 then $\operatorname{diam} \operatorname{supp} f = O(N)$, and this cannot be improved for functions f arising from (6.1). We show here [21] that we can choose the lattices Λ_j so that a common tiling function exists which is much more tight than that, tighter even than what Theorem 6 imposes. Of course our lattices will not satisfy the genericity condition (6.4) of Theorem 6, but will satisfy a lot of relations (their intersection will be a large lattice, in terms of density).

Fix a large prime p and consider the group $\mathbb{Z}_p^d$. Any nonzero element g of this group generates a cyclic subgroup of order p. It follows that $\mathbb{Z}_p^d$ has

$$\frac{p^d - 1}{p - 1} \sim p^{d-1} =: N$$

different cyclic subgroups. For each such subgroup G, which we now view as a subset of $\{0, 1, \ldots, p-1\}^d$, consider the lattice

$$\Lambda_G = (p\mathbb{Z})^d + G.$$

This contains the lattice $\Lambda = (p\mathbb{Z})^d$ and has volume

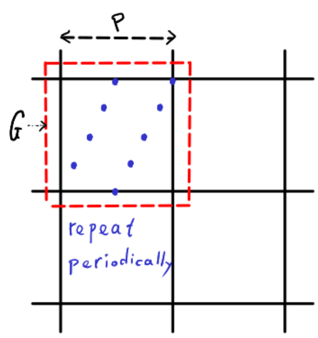


Figure 10: How we construct the $(p\mathbb{Z})^d$-periodic set Λ_G from the subgroup G of $\mathbb{Z}_p^d$

$$\operatorname{vol} \Lambda_G = \frac{\operatorname{vol}(p\mathbb{Z})^d}{|G|} = p^{d-1} = N.$$

The function $f = \mathbf{1}_{[0,p)^d}$, $[0,p)^d$ being a fundamental domain of Λ, tiles with Λ and, therefore, with any larger group, so f is a common tile of all Λ_G.

In order to make the volume of the Λ_G equal to 1 we shrink everything by $N^{1/d}$:

$$\Lambda'_G = N^{-1/d}\Lambda_G, \quad f'(x) = f(N^{1/d}x).$$

So we have $\sim N$ lattices Λ'_G of volume 1 and a common tile f' for them with

$$\operatorname{diam}\operatorname{supp} f' = \operatorname{diam}\operatorname{supp} f \cdot N^{-1/d} = \sqrt{d}\, p N^{-1/d} = \sqrt{d}\, N^{\frac{1}{d-1}-\frac{1}{d}} = \sqrt{d}\, N^{\frac{1}{d(d-1)}}.$$

We have proved:

Theorem 6. *In dimension $d \ge 2$ and for arbitrarily large N we can find N lattices of volume 1 and a common tile f for them with*

$$\operatorname{diam}\operatorname{supp} f = O_d\left(N^{\frac{1}{d(d-1)}}\right),$$

and, consequently, with

$$|\operatorname{supp} f| = O_d\left(N^{\frac{1}{d-1}}\right). \tag{10.1}$$

Question 6. *Derive a lower bound for* $\operatorname{diam}\operatorname{supp} f$*, for f tiling with $\Lambda_1, \ldots, \Lambda_N \subseteq \mathbb{R}^d$ and with $f \ge 0$ (or just $\int f > 0$) under no algebraic conditions for the lattices Λ_j, assuming only that* $\operatorname{vol}\Lambda_j \sim 1$.

Question 7. *In Theorem 10 we have used the cyclic subgroups of $\mathbb{Z}_p^d$ because they are easier to count. However the same argument could be carried out using a larger class of subgroups, perhaps all of them. What is the estimate that can be achieved this way to replace* (10.1)*?*

11 Almost matching upper and lower bounds for the diameter, $d = 1$

The construction that we used to prove Theorem 10 gives nothing in dimension $d = 1$. Yet, we can prove [21] that, if we allow relations among the lattices, we can achieve $\operatorname{diam}\operatorname{supp} f = o(N)$ in dimension 1 as well.

Let us start by defining

$$\lambda_j = \frac{1}{N+j}, \quad \Lambda_j = \lambda_j \mathbb{Z}, \quad (j = 1, 2, \ldots, N).$$

We will first construct a function f which tiles with all the Λ_j, $j = 1, 2, \ldots, N$, such that

$$\operatorname{diam}\operatorname{supp} f = o(1).$$

The Fourier transform of such an f must vanish on the dual lattices

$$\Lambda_j^* = \lambda_j^{-1}\mathbb{Z} = (N+j)\mathbb{Z}, \quad (j = 1, 2, \ldots, N)$$

except at 0. Write

$$U = \bigcup_{j=1}^{N} (N+j)\mathbb{Z} \ \setminus \{0\}.$$

By a result of Erdős [6] U, the set of integers which are divisible by one of the integers in $\{N+1, N+2, \ldots, 2N\}$, has density tending to 0 with N. Tenenbaum [28] has given the estimate that this density is at most

$$\frac{1}{\log^{\delta-o(1)} N}, \tag{11.1}$$

where $\delta = 0.086071\cdots$ is an explicit constant.

It is an important result of Beurling [2] that if Λ is a uniformly discrete set of real numbers of upper density ρ then for any $\epsilon > 0$ we can find a continuous function f, not identically zero, supported by the interval $[0, \rho+\epsilon]$ such that $\widehat{f}(\lambda) = 0$ for all $\lambda \in \Lambda$. We can even ask that $\widehat{f}(0) = 1$ if $0 \notin \Lambda$. By Tenenbaum's estimate (11.1) we can take $\rho = \log^{-\delta+o(1)} N$ and the set U, being a set of integers and thus uniformly discrete, satisfies the assumptions of Beurling's theorem, so there is a function f supported in the interval $[0, \log^{-\delta+o(1)} N]$, with integral 1, such that $\widehat{f} = 0$ on U. It follows that f tiles with all Λ_j.

We now scale by a factor of N

$$f'(x) = f(x/N), \quad \Lambda_j' = N\Lambda_j, \quad \operatorname{diam}\operatorname{supp} f' = O(N \log^{-\delta+o(1)} N)$$

and obtain the first half of the following theorem.

Theorem 7. *We can find N lattices $\Lambda_j \subseteq \mathbb{R}$ of with $\operatorname{vol}\Lambda_j \sim 1$ and a function f with $\int f > 0$ and supported in an interval of length*

$$\frac{N}{\log^{\delta-o(1)} N)}$$

which tiles with all Λ_j.

Furthermore, for any $\epsilon > 0$ any such function f must have

$$\operatorname{diam}\operatorname{supp} f \gtrsim_\epsilon N^{1-\epsilon}.$$

Arguing similarly we can also prove the lower bound for $\operatorname{diam}\operatorname{supp} f$ in Theorem 11. If we assume that f tiles with all $\Lambda_j = \lambda_j\mathbb{Z}$, with, say, $1 \le \lambda_j \le 2$, $j = 1, 2, \ldots, N$, then $\widehat{f}$ vanishes on

$$\bigcup_{j=1}^{N} \lambda_j^{-1}\mathbb{Z} \ \setminus \{0\}.$$

If this set is large then Jensen's formula implies that $\operatorname{diam}\operatorname{supp} f$ is also large. It was proved in [8, Theorem 1.1, special case $\ell = n$] that, for any $\epsilon > 0$, the above union of arithmetic progressions contains at least $c_\epsilon N^{2-\epsilon}$ points in $[0, 2N]$. By Jensen's formula then we have $\operatorname{diam}\operatorname{supp} f \gtrsim_\epsilon N^{1-\epsilon}$ and this completes the proof of Theorem 11.

Question 8. *Can we ensure $f \ge 0$ in the first half of Theorem 11?*

References

[1] J. Beck, *On a lattice point problem of H. Steinhaus*, Studia Sci. Math. Hung 24 (1989), 263–268.

[2] A Beurling, *Collected Works of Arne Beurling* (2 vol.), edited by L. Carleson et al, 1989.

[3] W. K. Chan and R. Mauldin, *Steinhaus tiling problem and integral quadratic forms*, Proceedings of the American Mathematical Society 135 (2007), no. 2, 337–342.

[4] M. Ciucu, *A remark on sets having the Steinhaus property*, Combinatorica 16 (1996), no. 3, 321–324.

[5] H. T. Croft, *Three lattice-point problems of Steinhaus*, The Quarterly Journal of Mathematics 33 (1982), no. 1, 71–83.

[6] Paul Erdős, *Note on sequences of integers no one of which is divisible by any other*, J. London Math. Soc 10 (1935), no. 1, 126–128.

[7] R. Gardner, *The Brunn-Minkowski inequality*, Bulletin of the American Mathematical Society 39 (2002), no. 3, 355–405.

[8] S. Gilboa and R. Pinchasi, *On the union of arithmetic progressions*, SIAM Journal on Discrete Mathematics 28 (2014), no. 3, 1062–1073.

[9] V. Havin and B. Jöricke, *The Uncertainty Principle in Harmonic Analysis*, Springer, Berlin, 1994.

[10] S. Jackson and R. Mauldin, *On a lattice problem of H. Steinhaus*, Journal of the American Mathematical Society 15 (2002), no. 4, 817–856.

[11] S. Jackson and R. D. Mauldin, *Sets meeting isometric copies of the lattice Z2 in exactly one point*, Proceedings of the National Academy of Sciences 99 (2002), no. 25, 15883–15887.

[12] S. Jackson and R. D. Mauldin, *Survey of the Steinhaus tiling problem*, Bulletin of Symbolic Logic 9 (2003), no. 03, 335–361.

[13] M. N. Kolountzakis, *A new estimate for a problem of Steinhaus*, International Mathematics Research Notices 1996 (1996), no. 11, 547–555.

[14] M. N. Kolountzakis, *A problem of Steinhaus: Can all placements of a planar set contain exactly one lattice point?*, Progress in Mathematics 139 (1996), 559–566.

[15] M. N. Kolountzakis, *Multi-lattice tiles*, International Mathematics Research Notices 1997 (1997), no. 19, 937–952.

[16] M. N. Kolountzakis, *The study of translational tiling with Fourier Analysis*, Fourier Analysis and Convexity (L. Brandolini, ed.), Birkhauser, 2004, pp. 131–187.

[17] M. N. Kolountzakis and M. Papadimitrakis, *The Steinhaus tiling problem and the range of certain quadratic forms*, Illinois Journal of Mathematics 46 (2002), no. 3, 947–951.

[18] M. N. Kolountzakis and M. Papadimitrakis, *Measurable steinhaus sets do not exist for finite sets or the integers in the plane*, Bulletin of the London Mathematical Society 49 (2017), no. 5, 798–805.

[19] M. N. Kolountzakis and T. Wolff, *On the Steinhaus tiling problem*, Mathematika 46 (1999), no. 02, 253–280.

[20] M. N. Kolountzakis and N. Lev, *Tiling by translates of a function: results and open problems*, Discrete Analysis (2021), 28122.

[21] M. N. Kolountzakis and E. Papageorgiou, *Functions tiling with several lattices*, arXiv preprint arXiv:2106.11701 (2021).

[22] P. Komjáth, *A lattice-point problem of Steinhaus*, The Quarterly Journal of Mathematics 43 (1992), no. 2, 235–241.

[23] E. Mallinikova and S. Rukshin, *On one Steinhaus problem*, Vestnik St. Petersburg Univ. Math. 28 (1995), no. 1, 28–32.

[24] R. Mauldin and A. Yingst, *Comments about the Steinhaus tiling problem*, Proceedings of the American Mathematical Society 131 (2003), no. 7, 2071–2079.

[25] R. D. Mauldin, *Some problems in set theory*, analysis and geometry, Paul Erdős and his Mathematics I (2001), 493–505.

[26] W. Sierpiński, *Sur un probleme de H. Steinhaus concernant les ensembles de points sur le plan*, Fundamenta Mathematicae 2 (1958), no. 46, 191–194.

[27] S. M. Srivastava and R. Thangadurai, *On Steinhaus sets*, Expositiones Mathematicae 23 (2005), no. 2, 171–177.

[28] G. Tenenbaum, *Lois de répartition des diviseurs*, Séminaire Delange-Pisot-Poitou. Théorie des nombres 19 (1980), no. 1, 1–3.

Mihail N. Kolountzakis: Department of Mathematics and Applied Mathematics, University of Crete, Voutes Campus, 70013 Heraklion, Crete, Greece.

E-mail: kolount@gmail.com

Prime Orbits and Möbius Randomness

Giorgos Kotsovolis

In memory of Dimitris Gatzouras

1 Introduction

The celebrated Prime Number Theorem was proved in 1896 independently by both Hadamard and de la Vallée Poussin. It constituted a major breakthrough in the field of number theory. It states that the number of prime numbers up to some integer N is asymptotically equal to $\frac{N}{\log N}$. If we denote by $\pi(N)$ the number of prime numbers $\leq N$, then it states

$$\pi(N) = \int_0^N \frac{dt}{\log t} + O\left(Ne^{-c\sqrt{\log N}}\right),$$

for some absolute constant c. (Note that throughout this transcript the Big O notation $f(x) = O(g(x))$ means that there is some constant C such that $-Cg(x) \leq f(x) \leq Cg(x)$.) Improving the error term of this asymptotic formula has proven to be one of the biggest obstacles in all of mathematics and it is related to the notorious Riemann Hypothesis. In particular, the Riemann Hypothesis is equivalent to the statement that for every $\varepsilon > 0$,

$$\pi(N) = \int_0^N \frac{dt}{\log t} + O\left(N^{\frac{1}{2}+\varepsilon}\right).$$

This is not the only interesting version of the Prime Number Theorem though. In the study of prime numbers one of the most important function is the Möbius function, which is the function $\mu : \mathbb{N} \to \{-1, 0, 1\}$ such that

$$\mu(n) = \begin{cases} 0, & \text{if } n \text{ is not square-free} \\ (-1)^r, & \text{if } n \text{ is square free} \end{cases},$$

where r is the number of primes appearing in the prime decomposition of n. An equivalent form of the Prime Number Theorem is the following equation concerning the Cesàro Averages of the Möbius function:

$$\frac{1}{N}\sum_{n=1}^{N} \mu(n) \to 0.$$

Again, the Riemann Hypothesis is related to the "speed" of that convergence, and it predicts that for every $\varepsilon > 0$,

$$\sum_{n=1}^{N} \mu(n) = O(N^{\frac{1}{2}+\varepsilon}).$$

Another theorem connected to the distribution of the prime numbers is Dirichlet's theorem regarding arithmetic progressions.

Theorem (Dirichlet). *For all pairs of integers (a, b) with $gcd(a, b) = 1$, the arithmetic progression $an + b$ contains infinitely many prime numbers.* Regarding the Möbius function, a closely related theorem states that

$$\frac{1}{N}\sum_{n=1}^{N}\mu(n)e^{2\pi ina} \to 0,$$

for all $a \in \mathbb{Q}$.

The purpose of this expository article is to explain the dynamical interpretations of these results and to generalize them in different settings, where one can hope to have generalized Prime Number Theorems.

2 Dynamical Interpretations

The aforementioned results carry dynamical interpretations. The function μ is a multiplicative function and one of the main principles of number theory is that additive structures and multiplicative structures behave independently. Say that two sequences a_n, b_n are **independent** if

$$\lim \frac{1}{N}\sum_{n=1}^{N} a_n b_n = \lim \frac{1}{N}\sum_{n=1}^{N} a_n \lim \frac{1}{N}\sum_{n=1}^{N} b_n.$$

It then happens that the equation $\frac{1}{N}\sum_{n=1}^{N}\mu(n)e^{2\pi ina} \to 0$ can be translated in that the Möbius sequence is independent from rational rotations. Hence, comes the question: What is an appropriate class of sequences that we can expect the Möbius function to be independent from? We need some definitions first (see [4]).

Definition. We say the the four-tuple $(\mathcal{X}, \mathcal{A}, \nu, T)$ is a **dynamical system** if $(\mathcal{X}, \mathcal{A}, \nu)$ is a measure space and $T : \mathcal{X} \to \mathcal{X}$ is a measure preserving function (i.e. for every $A \in \mathcal{A}$ we have that $\nu(T^{-1}(A)) = \nu(A)$).

The measure ν is called **ergodic** for the dynamical system $(\mathcal{X}, \mathcal{A}, \nu, T)$ iff for every set $A \in \mathcal{A}$ s.t. $T^{-1}(A) = A$ we have that $\nu(A) \in \{0, 1\}$. It probably seems like it is a lot to ask for a dynamical system to have this property, but it can easily be proved that if we consider the space $\mathcal{M}$ of measures ν that make T be measure preserving, then under the $weak^*$ topology, $\mathcal{X}$ is a compact, convex space with extreme points the ergodic measures. Hence, using functional analytic tools (such as the Krein-Milman Theorem), we see that most of the times assuming ergodicity is just a mild condition, simply because extreme points are the "building blocks" of a convex compact space.

The branch of mathematics that deals with the evolution of dynamical systems is called *Ergodic Theory*, and it has proven to be very closely related to Number Theory. Pick any function f in C($\mathcal{X}$)(i.e. the space of continuous functions assuming some topology on $\mathcal{X}$) and some $x \in \mathcal{X}$. Form the sequence $a_n(x) = f(T^n(x))$ (We will often be forgetting about the x-dependence). The nature of these sequences is additive (i.e. $a_{n+m}(x) = a_n(T^m(x))$) and one could expect that they are independent from the Möbius function. It has been proved that this is false. Every dynamical system can be associated with a non-negative number the **entropy** which is heuristically the measure of how random the system is in its evolution. In particular, when systems have zero entropy they are called **deterministic**, and when the systems have positive entropy they demonstrate random-like behaviour.

It has been proven that there are dynamical systems of positive entropy that have correlations with the Möbius function and hence cannot be orthogonal from it. However, for zero entropy systems there is a widely open conjecture due to P. Sarnak.

Möbius Disjointness Conjecture (P. Sarnak). *The Möbius function is orthogonal from every zero-entropy Dynamical System.*

This has proved to be an amazingly hard problem and only very few cases are known.

A very interesting class of dynamical systems are systems that come from Lie Groups, homogeneous spaces. The branch that deals with them is called *Homogeneous Dynamics* and is interesting from a number theoretical point of view.

Let G be some Lie Group and let Γ be a lattice that is a discrete subgroup such that the volume of a fundamental domain of G/Γ evaluated using the Haar measure of G is finite. If $a \in G$ we call a system of the form $(G/\Gamma, A, \mu_{haar}, x\Gamma \to gx\Gamma)$ a **homogeneous system**. Examples of these systems are the torus $\mathbb{R}/\mathbb{Z}$ or the space $SL_2(\mathbb{R})/SL_2(\mathbb{Z})$. These two systems are in some sense the prototypes, since for a random Lie Group G we have the two structural opposites that G can be nilpotent (like $\mathbb{R}$) and G can be semisimple (like $SL_2(\mathbb{R})$).

When a Homogeneous system originates from a nilpotent group G, it is called a *nilmanifold*, and nilmanifolds always have zero entropy. The Möbius Disjointness Conjeture was proven in 2008 in the case of nilmanifolds by B. Green and T. Tao [6], [5]. For semisimple homogeneous systems (i.e. when G is a semisimple Lie Group), the problem seems much harder and only a few cases are known. In 2011, J.Bourgain, P. Sarnak and T. Ziegler proved the first case for semisimple systems [1]. They proved Sarnak's conjecture for systems of the form $SL_2(\mathbb{R})/\Gamma$, where Γ is any lattice. In 2015, in his PhD thesis, R. Peckner generalized this result to higher dimensions, proving the conjecture for connected Real Groups, under a unipotent action (i.e. the element a in the definition of the translation of the homogeneous system is a unipotent element) [9].

3 Why should this conjecture hold?

Even though there have been some cases proved, why should one believe this conjecture? For one, P. Sarnak proved that the Chowla conjecture [2] would imply, if proved, the Möbius Disjointness Conjecture.

Chowla's Conjecture. *For any fixed integer m and exponents $a_1, a_2, \dots, a_m \geq 0$, with at least one of the a_i odd, we have*

$$\frac{1}{x}\sum_{n\leq x} \mu(n+1)^{a_1}\mu(n+2)^{a_2}\cdots\mu(n+m)^{a_m} \to 0.$$

Furthermore, a simpler version of Möbius Disjointness conjecture was proved in 2018, a result of B. Host and N. Fratzikinakis [3].

Theorem. *For every zero dynamical system $(\mathcal{X}, A, \nu, T)$ with zero entropy and at most countably many ergodic components (i.e. viewing ν as a point in the convex, compact space $\mathcal{M}$ of $T-$invariant measures, it can be written as an integral on countably many extreme points) and every $y \in \mathcal{X}$, we have that*

$$\frac{1}{\log N}\sum_{n=1}^{N}\frac{g(T^n(y))\mu(n)}{n} \to 0.$$

4 Why is Möbius Disjointness Important?

In one sentence, the Möbius Disjointness conjecture is the obstacle to allow us to have prime number theorems for dynamical systems and study the distribution of prime numbers in those systems.

Suppose we are given a dynamical system $(\mathcal{X}, \mathcal{A}, \nu, T)$, a function $f \in C(\mathcal{X})$. Birkhoff showed [4] that the sums of the form

$$\frac{1}{N}\sum_{n=1}^{N} f(T^n(x))$$

converge for ν-almost all points $x \in \mathcal{X}$. If the measure ν is an ergodic one, then it converges to $\int f(y)\, d\nu(y)$ for ν-almost all $x \in \mathcal{X}$. If, furthermore, the space M of invariant measures is a single set $\{\nu\}$, that is there exists a unique measure that makes T measure invariant, then the system is called *uniquely ergodic* and Birkhoff's result holds for all $x \in \mathcal{X}$.

A natural question to ask is the following: What if we restrict the summation only on prime numbers? Is a prime number theorem for the system $\mathcal{X}$ true? Will that also converge to $\int f(y)\, d\nu(y)$ even assuming that the system is uniquely ergodic? We restrict ourselves to 0 entropy systems, in order to have a chance to succeed. In mathematical terms, is it true that

$$\frac{\log N}{N}\sum_{p\leq N} f(T^p(x)) \to \int f(y)\, d\nu(y)?$$

We refer to results of this type as prime number theorems because for example if we take $\mathcal{X}$ to be the trivial 1-point system then this results is equivalent to the prime number theorem. If we look at the $1-$dimensional torus with a rational rotation this is equivalent with the prime number theorem on arithmetic progressions.

The next system we can look at is irrational rotations of the torus. In this case, Vinogradov managed to prove that the sequence ap_n is equidistributed, proving the conjecture and also providing an error term for the convergence. This proof is a part of his proof of the ternary Goldbach problem; which states that every big enough odd number is a sum of three primes. A key point of the proof is the notion of Type I and Type II sums.

5 Type I and Type I sums

For a dynamical system $(\mathcal{X}, \mathcal{A}, \nu, T)$ we call sums of the form

$$\frac{1}{N}\sum_{n=1}^{N} f(T^{kn}(x))$$

Type I sums, and sums of the form

$$\frac{1}{N}\sum_{n=1}^{N} f(T^{kn}(x))f(T^{ln}(x)),$$

Type II sums. Essentially thus, Type II sums are Type I sums for the system

$$(\mathcal{X}\times\mathcal{X}, \sigma(\mathcal{A}\times\mathcal{A}), \nu, T^k \times T^l).$$

In principle, good error terms for Type I and Type II sums can help prove the Möbius Disjointness Conjecture for the system $\mathcal{X}$ and a prime number theorem for the system.

B. Green and T. Tao in their series of papers managed in 2008 to prove that for all nilmanifolds $(G/\Gamma, \mathcal{A}, dg, x \to gx)$ one can get a logarithmic saving in Type I sums, meaning that assuming f is bounded by 1, $\frac{1}{N}\sum_{n=1}^{N} f(T^{kn})(x)$ can be bounded by $\frac{1}{(\log N)^c}$.

However, the product of two nilmanifolds is again a nilmanifold and thus this also holds for Type II sums. Hence, they proved the two conjectures for these spaces.

The situation is fundamentally different in semisimple systems. For example, as mentioned before, J. Bourgain, P. Sarnak and T. Ziegler proved Möbius Disjointness for $SL_2(\mathbb{R})/\Gamma$ under the action of the horocycle flow ($x \to \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} x$) and R. Peckner proved it for $SL_n(\mathbb{R})/\Gamma$. However, the distribution of prime orbits (i.e. the orbit restricted only on prime times) is still an open problem. Why?

Well, the answer is simply that we don't have good error terms. Actually, we do not have any error terms:

For example lets consider the example $SL_2(\mathbb{R})/\Gamma$ or even with $\Gamma = SL_2(\mathbb{Z})$. In their 2011 paper P. Sarnak and A. Ubis give good error terms for Type I sums in this case.

By computing effective equidistribution Theorems for Type I sums and using sieve methods they managed to prove that the closure of a generic prime orbit has volume at least $0,1$. That means that, at the very least the closure of the orbit under prime times is not too "weird", in the sense that it is not a null set. Of course, for a generic point x the expected result is that the prime orbit will be dense and even more, equidistributed.

However, Type II sums are a very different story and that is because the space $SL_2(\mathbb{R}) \times SL_2(\mathbb{R})$ is not that well understood. The only tools, so far, we have to study Type I sums on that space are Marina Ratner's Theorems [10], that can reduce Type II sums to the study of commensurator groups. At this points it should be mentioned that Elon Lindenstrauss and Amir Mohammadi managed to give an effective density theorem for type II sums on $SL_2(\mathbb{R})/\Gamma$ when Γ is an arithmetic lattice [8]. This gives hope for effective equidistribution results.

6 Horocycle flows

Among the homogeneous spaces one might try to understand in order to prove the aforementioned conjectures, horocycles flows are certainly one of the most interesting classes. We define what a horocycle flow is now. Suppose we have a homogeneous space G/Γ and a point $x \in G$. Define

$$U = \{a \in G, g^n a g^{-n} \to e\}.$$

U is called a **horosphere** and the reason for the name is that if $G = SL2(\mathbb{R})$ and g is a non-trivial point of the geodesic flow, such as $\begin{pmatrix} 2 & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$, the orbit Ux for $x \in G$ looks like a generalized circle (could be a line in some cases). Horocycle flows, that is one parameter families inside horospheres are examples of unipotent elements, as it can easily be noticed by the defining equations that all points in such a flow have eigenvalues 1, when treated as groups of matrices. Unipotent flows are of central role because of Ratner's theorems [10](See next section).

The action of such groups is also a great place to look at because they always have zero entropy. The dynamics of such spaces are particularly beautiful and better understood than general actions.

Take once more the example of $\mathcal{X} = SL_2(\mathbb{R})/SL_2(\mathbb{Z})$. In this space, the action of $u_t = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$ is ergodic and also mixing:

If H is a group acting on a space $L^2(\mathcal{X})$, we call the action *mixing* if for any $f, g \in L^2(\mathcal{X})$,

$$\int_{\mathcal{X}} f(u_t x) g(x)\, dg(x) \to \int_{\mathcal{X}} g(x)\, dg(x) \int_{\mathcal{X}} f(x)\, dg(x).$$

This is a much stronger property than ergodicity. Even further, this mixing convergence can be made effective and in fact it is known that the decay is at least exponential[7], [13], [12]:

$$\left| \int_{\mathcal{X}} f(u_t x) g(x)\, dg(x) - \int_{\mathcal{X}} g(x)\, dg(x) \int_{\mathcal{X}} f(x)\, dg(x) \right| \ll e^{-t\kappa}.$$

The decay is linked with the spectral gap of the space, that is the first eigenvalue of the Laplacian. Big questions related to this problem are the Selberg and Ramanujan Conjectures.

7 Unipotent Orbits and Ratner

In the case of $SL_2(\mathbb{R})$, two of the most interesting transformations one can consider to make the space into a dynamical system are the geodesic flow and the horocycle flow. The **geodesic flow** is the transformation $g_t : x \to \begin{pmatrix} e^{-t} & 0 \\ 0 & e^t \end{pmatrix} x$. However, this system has positive entropy. It is thus not to interest for us. The horocycle flow is the system with transformation $u_t : x \to \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} x$ This system does have zero entropy. Actions like that coming from unipotent elements are particularly beautiful as explained by the rigidity theorems of Marina Ratner, as in the following.

Theorem (Marina Ratner). *Let G/Γ a homogeneous spaces and u_t a one parameter subgroup consisting of unipotent element and $x \in G/\Gamma$. Then $u_t x$ is eduidistributed in its closure and its closure is algebraic in the sense that it is some homogeneous space H/Γ for some H subgroup of G containing the family $\{u_t\}$.*

Specifically for the case of $SL_2(\mathbb{R})/\Gamma$, we have that when Γ is cocompact every orbit is equidistributed (that is, we have unique ergodicity) and when Γ is not cocompact every orbit is either equidistributed in the whole space or isomorphic to a one dimensional torus. Which case occurs has to do with whether the orbit touches the real line on an irrational point or not, respectively.

To understand how much better the unipotent case is from the geodesic one, in the geodesic case the orbits can be dense but very far from equidistributed or even worse form very complicated objects.

Unfortunately, effectivization of Ratner's proof is a very hard task and not much is known around that problem [8].

As mentioned earlier, P. Sarnak and A. Ubis [11] gave an effective version of Ratner's theorem for the space $SL_2(\mathbb{R})$ explaining Type I sums for these spaces. However, there is no effective Ratner theorem for $SL_2(\mathbb{R}) \times SL_2(\mathbb{R})$ yet.

It has been proved that most Type II sums are well behaved converging to $\left(\int_{G/\Gamma} f(y)dg\right)^2$, however nothing is known for the speed of that convergence, forbidding us to pass to prime orbits.

References

[1] J. Bourgain, P. Sarnak and T. Ziegler, *Disjointness of Möbius from horocycle flows*, From Fourier analysis and number theory to Radon transforms and geometry, Dev. Math., vol. 28, Springer, New York, 2013, pp. 67–83.

[2] S. Chowla, *The Riemann hypothesis and Hilbert's tenth problem*, Mathematics and its Applications, Vol. 4 Gordon and Breach Science Publishers, New York-London-Paris 1965 xv+119 pp.

[3] N. Frantzikinakis and B. Host, *The logarithmic Sarnak conjecture for ergodic weights*, Ann. of Math. (2) **187** (2018), no. 3, 869–931.

[4] M. Einsiedler and T. Ward, *Ergodic theory with a view towards number theory*, Graduate Texts in Mathematics, 259. Springer-Verlag London, Ltd., London, 2011. xviii+481 pp.

[5] B. Green and T. Tao, *The Möbius function is strongly orthogonal to nilsequences*, Ann. of Math. (2) **175** (2012), no. 2, 541–566.

[6] B. Green and T. Tao, *The quantitative behaviour of polynomial orbits on nilmanifolds*, Ann. of Math. (2) **175** (2012), no. 2, 465–540.

[7] D. Y. Kleinbock and G. A. Margulis, *Bounded orbits of nonquasiunipotent flows on homogeneous spaces*, Sinai's Moscow Seminar on Dynamical Systems, Amer. Math. Soc. Transl. Ser. 2, vol. 171, Amer. Math. Soc., Providence, RI, 1996, pp. 141–172.

[8] E. Lindenstrauss and A. Mohammadi, *Polynomial effective density in quotients of* $\mathbb{H}^3$ *and* $\mathbb{H}^2 \times \mathbb{H}^2$, Preprint.

[9] R. Peckner, *Möbius disjointness for homogeneous dynamics*, Duke Math. J. **167** (2018), no. 14, 2745–2792.

[10] M. Ratner, *On Raghunathan's measure conjecture*, Ann. of Math. (2) **134** (1991), no. 3, 545–607.

[11] P. Sarnak and A. Ubis, *The horocycle flow at prime times*, J. Math. Pures Appl. (9) **103** (2015), no. 2, 575–618.

[12] A. Strömbergsson, *On the deviation of ergodic averages for horocycle flows*, J. Mod. Dyn. **7** (2013), no. 2, 291–328.

[13] A. Venkatesh, *Sparse equidistribution problems, period bounds, and subconvexity*, Ann. of Math. (2) **172** (2010), no. 2, 989–1094.

Giorgos Kotsovolis: Princeton University, Department of Mathematics, Fine Hall, Washington Road Princeton NJ 08544-1000 USA.

E-mail: gk13@princeton.edu

From Compressed Sensing to Deep Unfolding: Redundancy Unites

Vicky Kouni

In Memory of D. Gatzouras

Abstract

We examine Compressed Sensing from a model-based and a data-driven point of view. In the first part, we solve analysis-sparsity-based Compressed Sensing (analysis CS), employing spark deficient Gabor frames, and compare numerically our method with state-of-the-art Gabor transforms. Our results confirm that the high redundancy provided by spark deficient Gabor frames improves the performance and reconstruction quality of the CS algorithm. In the second part, we propose a new deep unfolding network coined DECONET, which jointly learns a redundant sparsifying analysis operator and solves the analysis CS problem. We deliver meaningful – in terms of sparsifier's redundancy and number of layers – generalization error bounds for DECONET, using a chaining technique. Finally, we confirm the validity of our theoretical results in terms of adequate numerical experiments.

1 Model-based Compressed Sensing

1.1 Introduction to CS & Sparse Data Models

Compressed Sensing (CS) deals with recovering $x \in \mathbb{F}^n$ ($\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$) from incomplete, noisy measurements $y = Ax + e \in \mathbb{F}^m$, $m < n$, $\|e\|_2 \leq \varepsilon$. CS heavily relies on the following two principles. First, the measurement matrix A must meet conditions (small coherence, restricted isometry property [1], etc.) ensuring exact/approximate reconstruction of x. Random Gaussian matrices have proven to work well. Second, we impose a *sparse data model* for x [2], that is, we assume that x has very few non-zero coefficients.

Sparse data models are split in synthesis and analysis sparsity. In the former [1, 2], x is considered to be sparse when $x = Ds$. $D \in \mathbb{F}^{n\times n}$ is an orthogonal/unitary matrix and s – the *synthesis representation* of x – is sparse. On the other hand, in the analysis sparsity model [3, 4, 5], we assume there exists a *redundant analysis operator* $W \in \mathbb{R}^{N\times n}$, $N > n$, so that Wx is sparse. The two models coincide when $W = D^{-1}$. But the interesting case is when considering a redundant W, so analysis sparsity differentiates itself from its "twin". In this paper, we solely focus on analysis sparsity. The optimization problem that emerges when employing analysis sparsity in CS is the *analysis l_1-minimization* problem

$$\min_{x\in\mathbb{F}^n} \|Wx\|_1 \quad \text{subject to} \quad \|Ax - y\|_2 \leq \varepsilon. \tag{1.1}$$

As mentioned earlier, we prefer to employ analysis sparsity in CS, due to some advantages it has compared to its synthesis counterpart. For example, analysis sparsity provides flexibility in modelling sparse signals, since it leverages the redundancy of the involved analysis operators [6]. Moreover, it is computationally more appealing to solve the optimization algorithm of analysis CS since a) the actual optimization takes place in the ambient space b) the algorithm may need less measurements for perfect reconstruction [3].

1.1.1 Related Work, Motivation & Contributions

We are inspired by [3, 4], which propose analysis operators associated with *full spark frames* (=overcomplete bases) [7]. However, the full spark property is not beneficial for analysis l_1-minimization, since it restricts the amount of linear dependencies in W, which in turn could be leveraged by (1.1). To overcome this obstacle, we need analysis operators exhibiting more linear dependencies among their rows. To that end, we employ *spark deficient Gabor frames* [8], which have enhanced linear dependencies and are little explored in terms of analysis CS. We associate to such a frame a redundant analysis operator W. In the end, we present numerical comparisons of the proposed operator to 3 famous Gabor analysis operators and explore an interesting number-theoretic outcome.

1.2 Spark Deficient Gabor Frames

Definition. A *discrete Gabor system* (g, a, b) is defined as a collection of time-frequency shifts of the so-called window vector $g \in \mathbb{C}^n$, expressed as

$$g_{p,q}(k) = e^{2\pi iqbk/n} g(k - pa), \quad k = 0, \dots n - 1, \tag{1.2}$$

a, b denoting time and frequency (lattice) parameters respectively, $p = 0, \dots, n/a - 1$ ($n/a \in \mathbb{N}$) and $q = 0, \dots, n/b - 1$ ($n/b \in \mathbb{N}$) denoting time and frequency shift indices, respectively. If (g, a, b) with elements as in (1.2) spans $\mathbb{C}^n$, then it is a *Gabor frame* for $\mathbb{C}^n$ and $ab < n$.

Remark 4. Due to (1.2), the number of elements in (g, a, b) is $N = n^2/ab$. Furthermore, a, b play a crucial role in a signal's good time-frequency resolution w.r.t. a Gabor frame, but they are typically chosen through experiments.

Definition (Gabor Transform)**.** $W_g : \mathbb{C}^n \mapsto \mathbb{C}^N$ denotes the *Gabor analysis operator* – also known as *digital Gabor transform* (DGT) – whose action on a signal $x \in \mathbb{C}^n$ is defined as

$$W_g : x \mapsto W_g x = (\langle x, g_l \rangle)_{l=0}^{N-1}. \tag{1.3}$$

Definition (Spark of a frame in $\mathbb{C}^n$)**.** The *spark* of a set Φ –denoted by $\mathrm{sp}(\Phi)$– of N vectors in $\mathbb{C}^n$ is the size of the smallest linearly dependent subset of Φ. A frame Φ is *full spark* if and only if every set of n elements of Φ is a basis, or equivalently $\mathrm{sp}(\Phi) = n + 1$, otherwise it is *spark deficient*.

We state here the following auxiliary definition.

Definition. The symplectic group $\mathrm{SL}(2, \mathbb{Z}_n)$ consists of all matrices

$$G = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \tag{1.4}$$

such that α, β, γ, $\delta \in \mathbb{Z}_n$ and $\alpha\delta - \beta\gamma \equiv 1(\mathrm{mod} n)$. To each such matrix corresponds a unitary matrix given by the explicit formula

$$U_G = \frac{\exp(i\theta)}{\sqrt{n}} \sum_{u,v=1}^{n} \tau^{\beta^{-1}(\alpha(v-1)^2 - 2(u-1)(v-1) + \delta(u-1)^2)} e_u e_v, \tag{1.5}$$

where θ is an arbitrary phase, $\beta\beta^{-1} \equiv 1 \mathrm{mod} n$ and $\tau = -\exp(\frac{i\pi}{n})$.

We restrict ourselves to a specific $\mathcal{Z} \in \mathrm{SL}(2, \mathbb{Z}_n)$, that is, the *Zauner* matrix

$$\mathcal{Z} = \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix} \equiv \begin{pmatrix} 0 & n-1 \\ 1 & n-1 \end{pmatrix}, \tag{1.6}$$

with corresponding unitary $U_{\mathcal{Z}}$ (produced by combining (1.5) and (1.6)).

Theorem 1.1 ([9]). *Let $n \in \mathbb{N}$ such that $2 \nmid n$, $3 \mid n$ and n is square-free. Then, any eigenvector of the Zauner unitary matrix $U_{\mathcal{Z}}$, generates a spark deficient Gabor frame for $\mathbb{C}^n$.*

Remark 5. A simple way to choose n relies on prime factorization: take k prime numbers $p_1^{\alpha_1}, \dots, p_k^{\alpha_k}$, with $\alpha_1, \dots, \alpha_k$ not all a multiple of 2 and $p_1 = 3, p_i \neq 2, i = 2, \dots, k$, such that $n = 3^{\alpha_1} p_2^{\alpha_2} \cdot \dots \cdot p_k^{\alpha_k}$. Since $a, b \mid n$, we may also choose $a = 1$ and $b = p_i^{\alpha_i}, i = 1, \dots, k$. Otherwise, both a, b can be taken from the prime factorization of n. We have seen empirically that this method for fixing (n, a, b) produces satisfying results.

1.3 Numerical Experiments

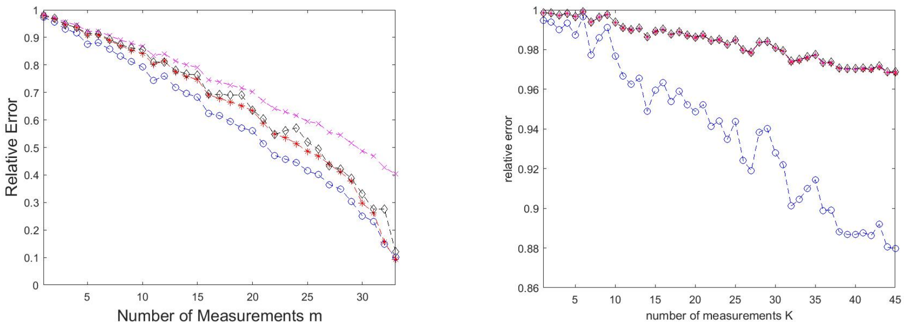
We compare star-DGT to 3 other DGTs generated by state-of-the-art window vectors (Gaussian, Hann, Hamming) in time-frequency analysis on:

- 2 synthetic real-valued signals: Cusp & Sing (Wavelab),
- 2 real-world speech signals (TIMIT),

by solving four different instances of (1.1) (one for each DGT) and report the results in Fig. 1 and Table 1.

1.3.1 Discussion

As illustrated in Fig. 1, star-DGT outperforms all DGTs based on state-of-the-art window vectors. Furthermore, we observe in Table 1 that for different pairs of (a, b) for a given $x \in \mathbb{R}^n$, star-DGT outperforms the rest of DGTs. Interestingly, all DGTs achieve the smallest relative error when a, b are chosen as the two largest factors in the prime factorization of n. Overall, our results indicate improved performance of analysis l_1-minimization when employing W_{g_*}.



(a) Cusp with $(L, a, b) = (33, 1, 11)$ (b) Sing with $(L, a, b) = (45, 1, 9)$

Figure 1: Rate of approximate success for synthetic signals, with a randomly subsampled identity $A \in \mathbb{R}^{m \times n}$ and additive zero-mean Gaussian noise with std= 0.001. Red: Gaussian, magenta: Hann, black: Hamming, blue: proposed.

	Relative Error $\|\hat{x} - x\|_2 / \|x\|_2$					
Parameters	SI2141, $n = 21735$			SX224, $n = 24633$		
Windows \ (a, b)	$(5, 7)$	$(9, 21)$	$(7, 23)$	$(7, 17)$	$(17, 23)$	$(21, 23)$
Gaussian	0.8916	0.8904	0.8897	0.8918	0.8910	0.8835
Hann	0.8917	0.8905	0.8897	0.8933	0.8925	0.8939
Hamming	0.8917	0.8905	0.8897	0.8932	0.8923	0.8940
star	$\mathbf{0.8759}$	$\mathbf{0.8743}$	$\mathit{0.8734}$	$\mathbf{0.8881}$	$\mathit{0.8828}$	$\mathbf{0.8835}$

Table 1: Fixed $m = \frac{n}{4}$ with varying (a, b), for two example speech signals. Bold letters indicate the best performance among all windows. Italic letters indicate the min. rel. error achieved by star window among all pairs (a, b).

1.4 Conclusion and Future Work

In this paper, We took advantage of a window vector to generate a spark deficient Gabor frame and associated to it a (highly) redundant analysis operator, namely star-DGT. We numerically compared star-DGT to 3 other DGTs generated by common window vectors in the field of Gabor Analysis, by solving the analysis l_1-minimization problem, for synthetic and real-world data. Our experiments confirm improved performance: the increased amount of linear dependencies provided by this SDGF, yields in all cases a lower relative reconstruction error, as m increases. As future direction, we would like to develop a mathematical framework explaining why star-DGT benefits more when (a, b) are chosen as the two largest primes in the prime factorization of n.

2 Deep Learning in Compressed Sensing

2.1 Introduction to Deep Unfolding Networks

Model-based CS has the advantage of being well-studied in the last 16 years, with many theoretical guarantees explaining its success. Nevertheless, the employed optimization algorithms typically need many iterations to converge to a solution of the problem and the quality of the reconstructed signal may be low, even for $m = n/2$. A plausible recent alternative relies on *Deep Neural Networks* (DNNs), which offer higher reconstruction quality in a significantly faster way.

Definition (DNN). A Deep Neural Network with L *layers* is a tuple

$$\Phi = ((T_1, \sigma_1), \ldots, (T_L, \sigma_L)), \tag{2.1}$$

where $T_\ell : \mathbb{R}^{N_{\ell-1}} \mapsto \mathbb{R}^{N_\ell}$, with $T_\ell(x) = W_\ell x + b_\ell$, are affine maps and $\sigma_\ell : \mathbb{R}^{N_\ell} \mapsto \mathbb{R}^{N_\ell}$, $\ell = 1, \ldots, L$, are (nonlinear) *activation* functions. For $\ell = 1, \ldots, L$, we call W_ℓ *weights*, b_ℓ *biases* and the output of the ℓth layer of Φ has the form $\sigma_\ell(W_\ell x + b_\ell)$. The function $h : \mathbb{R}^{N_0} \mapsto \mathbb{R}^{N_L}$ given by $h = \sigma_L \circ T_L \circ \cdots \circ \sigma_1 \circ T_1$ is the *realization* of Φ.

Definition. A *deep residual network* with L layers and shared parameters W, b is a DNN with the same W_ℓ, b_ℓ and σ_ℓ across all layers, i.e.,$W_\ell \equiv W, b_\ell \equiv b$, $\sigma_\ell \equiv \sigma$, $\forall \ell = 1, \ldots, L$. The ℓth layer of such a DNN has the form $\sigma(Wx + b) + x$.

The task of a DNN is the approximation of a function $f : \mathbb{R}^{N_0} \mapsto \mathbb{R}^{N_L}$, based on a given *training set* $\mathcal{S} = \{(x_i, y_i)\}_{i=1}^s = \{(x_i, f(x_i)\}_{i=1}^s$, drawn i.i.d. from an (unknown) distribution $\mathcal{D}^s$. In optimization language, this pertains to finding/*learning* the unknown parameters $\boldsymbol{m}\theta_\ell = (W_\ell, b_\ell)$ that minimize the training *mean-squared error* (MSE)

$$\mathrm{MSE}_{train}(h) = \frac{1}{s} \sum_{i=1}^{s} \| h_{(\boldsymbol{m}\theta_1, \ldots, \boldsymbol{m}\theta_L)}(x_i) - f(x_i) \|_2^2. \tag{2.2}$$

A new line of research lies on merging DNNs and optimization algorithms, leading to the so-called *deep unfolding/unrolling* [10]. The latter pertains to interpreting the iterations of well-studied optimization algorithms for analysis CS as layers of a DNN – called deep unfolding network (DUN) – which implements a *decoder*: given measurements $y \in \mathbb{R}^m$, we get $h(y) = \hat{x} \approx x \in \mathbb{R}^n$.

2.2 Relater Work, Motivation & Contribution

State-of-the-art (SotA) DUNs jointly learn a decoder for CS and a sparsifying transform [11], [12], [13]. Moreover, there exists few recent studies on the *generalization error* [14] of proposed unfolding networks [11]. The drawback of [11, 12] is that the learnable sparsifiers satisfy some orthogonality constraint, so CS is solved under the synthesis sparsity model. On the other hand, [13] employs its handier analysis counterpart; but it provides no generalization analysis of the proposed framework. Similarly, we propose a new DUN dubbed DECONET, based on an optimal analysis-l_1 algorithm [15]. DECONET jointly learns a decoder for CS and a *redundant sparsifying analysis operator*; thus, we address the CS problem under the analysis sparsity model. Our novelty lies on estimating the generalization error of the proposed analysis-based unfolding network. To that end, we upper bound the generalization error of DECONET in terms of the Rademacher complexity [16] of the associated hypothesis class. In the end, we numerically confirm the validity of our theoretical findings.

2.3 Derivation of a New Unfolding Network for CS

As stated in [15], the analysis l_1-minimization problem (1.1) transforms into

$$\min_{x\in\mathbb{R}^n} \|Wx\|_1 + \frac{\mu}{2}\|x-x_0\|_2^2 \quad \text{s. t.} \quad \|y-Ax\|_2 \leq \varepsilon, \tag{2.3}$$

where μ is a *smoothing* parameter and x_0 is an initial guess on x. Then, (2.3) is associated to the dual

$$\begin{aligned} \text{maximize} \quad & \langle y, z^2\rangle - \varepsilon\|z^2\|_2 \\ \text{s. t.} \quad & A^T z^2 - W^T z^1 = 0, \ \|z^1\|_\infty \leq 1. \end{aligned} \tag{2.4}$$

Finally, after a collection of arguments and calculations, Algorithm 1 is presented (we call it *analysis conic form* (ACF) from now on), with $\mathcal{S}(\cdot)$ and $\mathcal{T}(\cdot)$ being soft-thresholding and truncation operators, respectively. ACF also involves step sizes $\{t_k^1\}$, $\{t_k^2\} > 0$ and a step size multiplier $0 < \{\theta_k\}$, with typical update rules giving $0 < \{t_k^1\}_{k\geq 0}$, $\{t_k^2\}_{k\geq 0} \leq 1$, $0 < \{\theta_k\}_{k\geq 0} \leq 1$.

Algorithm 1: ACF

Input : $x_0 \in \mathbb{R}^n$, $z_0^1 \in \mathbb{R}^N$, $z_0^2 \in \mathbb{R}^m$, $\mu \in \mathbb{R}_+$, step sizes $\{t_k^1\}, \{t_k^2\}$

Output: solution $\hat{x}_\mu$ of (2.3)

$\theta_0 \leftarrow 1$, $u_0^1 = z_0^1$, $u_0^2 = z_0^2$;

for iterations $k = 0, 1, \ldots$ **do**

  $x_k \leftarrow x_0 + \mu^{-1}((1-\theta_k)W^T u_k^1 + \theta_k W^T z_k^1 - (1-\theta_k)A^T u_k^2 - \theta_k A^T z_k^2)$;

  $z_{k+1}^1 \leftarrow \mathcal{T}((1-\theta_k)u_k^1 + \theta_k z_k^1 - \theta_k^{-1} t_k^1 W x_k, \theta_k^{-1} t_k^1)$;

  $z_{k+1}^2 \leftarrow \mathcal{S}((1-\theta_k)u_k^2 + \theta_k z_k^2 - \theta_k^{-1} t_k^2 (y - Ax_k), \theta_k^{-1} t_k^2 \varepsilon)$;

  $u_{k+1}^1 \leftarrow (1-\theta_k)u_k^1 + \theta_k z_{k+1}^1$;

  $u_{k+1}^2 \leftarrow (1-\theta_k)u_k^2 + \theta_k z_{k+1}^2$;

  $\theta_{k+1} \leftarrow 2/(1 + (1 + 4/(\theta_k)^2)^{1/2})$;

end for

Remark 6. According to [15], ACF is guaranteed to converge to a solution $\hat{x}_\mu$ of (2.3), for which $\hat{x}_\mu \overset{\mu\to 0}{\longrightarrow} \hat{x}$, with $\hat{x}$ being an optimal solution of (1.1). Furthermore, the optimal solution $\hat{x}$ is identified to this uniquely determined $\hat{x}_\mu$ and it is argued that there are situations where $\hat{x}$ and $\hat{x}_\mu$ coincide. Henceforward, we stick to the formulation of [15] and speak about the solution $\hat{x}$.

We consider a standard scenario for ACF with $z_0^1 = u_0^1 = 0$, $z_0^2 = u_0^2 = 0$, $t_0^1 = t_0^2 = \theta_0 = 1$, $0 < \{t_k^1\}, \{t_k^2\}, \{\theta_k\} \leq 1$, $\mu > 1$, $x_0 = A^T y$, with $A \in \mathbb{R}^{m\times n}$ being an appropriately normalized random matrix, with $\|A\|_{2\to 2} \approx 1$. We concatenate the dual variables in one vector, i.e., $v_k^T = (z_k^1, z_k^2, u_k^1, u_k^2)^T \in \mathbb{R}^{1\times(2N+2m)}$ for $k \geq 0$, with $v_0 = 0$, and do the calculations, so that

$$v_{k+1} = D_k v_k + \Theta_k \begin{pmatrix} \mathcal{T}(G_k^1 v_k - b_k^1, \theta_k^{-1} t_k^1) \\ \mathcal{S}(G_k^2 v_k - b_k^2, \theta_k^{-1} t_k^2 \varepsilon) \\ \mathcal{T}(G_k^1 v_k - b_k^1, \theta_k^{-1} t_k^1) \\ \mathcal{S}(G_k^2 v_k - b_k^2, \theta_k^{-1} t_k^2 \varepsilon) \end{pmatrix}, \quad k \geq 0, \tag{2.5}$$

where D_k, $\Theta_k \in \mathbb{R}^{(2N+2m)\times(2N+2m)}$ depend on θ_0, θ_k; $G_k^1 \in \mathbb{R}^{N\times(2N+2m)}$ depends on W, A, t_k^1, θ_k, μ; $G_k^2 \in \mathbb{R}^{m\times(2N+2m)}$ depends on W, A, t_k^2, θ_k, μ; $b_k^1 \in \mathbb{R}^N$ depends on W, t_k^1, θ_k, x_0; $b_k^2 \in \mathbb{R}^m$ depends on A, t_k^2, θ_k, x_0 and y. As one may notice, (2.5) resembles a DNN. In order to enable learning, we assume that W is unknown and learned from a training sequence $\mathcal{S} = \{(x_i, y_i)\}_{i=1}^s \overset{i.i.d.}{\sim} \mathcal{D}$, with unknown $\mathcal{D}^s$. Hence, the trainable parameters are the entries of W. Moreover, we consider W to be bounded with respect to the operator norm $\|\cdot\|_{2\to 2}$ by some $\Lambda > 0$, so we write $W \in \mathcal{B}_\Lambda$.

Based on (2.5), we formulate ACF as a neural network with L layers/iterations and output of the kth layer defined as

$$f_1(y) = \sigma(y), \tag{2.6}$$

$$f_k(v) = D_{k-1}v + \Theta_{k-1}\sigma(v), \quad k = 2, \dots, L, \tag{2.7}$$

where

$$\begin{aligned}\sigma(y)^T &= (\mathcal{T}(-t_0^1 W x_0, t_0^1), \mathcal{S}(t_0^2(y - Ax_0), t_0^2\varepsilon),\\ &\qquad \mathcal{T}(-t_0^1 W x_0, t_0^1), \mathcal{S}(t_0^2(y - Ax_0), t_0^2\varepsilon))^T, \end{aligned}\tag{2.8}$$

$$\begin{aligned}\sigma(v)^T &= (\mathcal{T}(G_k^1 v - b_k^1), \mathcal{S}(G_k^2 v - b_k^2),\\ &\qquad \mathcal{T}(G_k^1 v - b_k^1), \mathcal{S}(G_k^2 v - b_k^2))^T, \quad k = 2, \dots, L.\end{aligned}\tag{2.9}$$

We denote the composition of L such layers (all having the same W) as

$$f_W^L(y) = f_L \circ f_{L-1} \circ \cdots \circ f_1(y). \tag{2.10}$$

The *intermediate decoder* of (2.10) constitutes the realization of a DNN with L layers, that reconstructs v from y. Now, we apply

- an affine $\phi : \mathbb{R}^{(2N+2m)\times 1} \mapsto \mathbb{R}^n$, so that $\hat{x} := \phi(f_W^L(y))$,
- a truncating $\psi : \mathbb{R}^n \to \mathbb{R}^n$ to push $\phi(f_W^L(y))$ inside a l_2-ball of radius B_{out}, for some constant $B_{\text{out}} > 0$.

For a fixed number of layers L, the learnable decoder is

$$\mathrm{dec}_W^L(y) = \psi(\phi(f_W^L(y))). \tag{2.11}$$

We call *Decoding Network* (*DECONET*) the DNN that implements such a decoder, which is parameterized by W.
We introduce the hypothesis class of DECONET

$$\mathcal{H}^L = \{h : \mathbb{R}^m \mapsto \mathbb{R}^n : h(y) = \psi(\phi(f_W^L(y))),\, W \in \mathcal{B}_\Lambda\}, \tag{2.12}$$

consisting of all the functions/decoders that DECONET can implement. Given $\mathcal{H}^L$ and $\mathcal{S}$ (with $|\mathcal{S}| = s$), DECONET yields a decoder $h_{\mathcal{S}} \in \mathcal{H}^L$ that aims at reconstructing x from y. In order to measure the reconstruction error of a hypothesis $h \in \mathcal{H}^L$ on $\mathcal{S}$, we employ the train MSE (2.2). The *true loss* is

$$\mathcal{L}(h) = \mathbb{E}_{(x,y)\sim\mathcal{D}}(\|h(y) - x\|_2^2). \tag{2.13}$$

We are interested in the generalization error, given as the difference between the train MSE and the true loss

$$\mathrm{GE}(h) = |\hat{\mathcal{L}}_{train}(h) - \mathcal{L}(h)|. \tag{2.14}$$

The generalization error is important, because it tells us how well a DNN performs on unseen data. But since $\mathcal{D}$ is unknown, wwe estimate (2.14) in terms of the *empirical Rademacher complexity* [16] associated to $\mathcal{H}^L$:

$$\mathcal{R}_{\mathcal{S}}(\mathcal{H}^L) = \mathbb{E} \sup_{h \in \mathcal{H}^L} \sum_{i=1}^{s} \sum_{k=1}^{n} \epsilon_{ik} h_k(x_i), \tag{2.15}$$

ϵ being a Rademacher vector. We do so in a series of steps, described in the following subsections. The proofs of all our results can be found in [17].

2.4 Theoretical Results

2.4.1 Boundedness of DECONET's Outputs

We make some typical (for the machine learning literature) assuptions for $\mathcal{S}$, i.e., we assume that $\|x\|_2 \leq B_{\text{in}}$ for some constant $B_{\text{in}} > 0$ and $\|h(y)\|_2 \leq B_{\text{out}}$ for some constant $B_{\text{out}} > 0$; thus, $\|h(y) - x\|_2 \leq \|x\|_2 + \|h(y)\|_2 \leq B_{\text{in}} + B_{\text{out}}$. We now take into account the number of training samples and pass to matrix notation. Hence, $\|Y\|_F \leq \sqrt{s}\mathrm{B}_{\text{in}}$ and

$$\|h(Y)\|_F = \|\psi(\phi(f_W^L(Y)))\|_F \leq \sqrt{s}\mathrm{B}_{\text{out}}. \tag{2.16}$$

We upper-bound the output $f_W^k(Y)$ with respect to the Frobenius norm, after any number of layers k and especially for $k < L$, so that ϕ and ψ are not applied after the final layer L.

Lemma 2.1. *Let $k \in \mathbb{N}$. For any $W \in \mathcal{B}_\Lambda$, step sizes $0 < \{t_k^1\}, \{t_k^2\} \leq 1$ with $t_0^1 = t_0^2 = 1$, $t_{-1}^1 = t_{-1}^2 = 0$, step size multiplier $0 < \{\theta_k\} \leq 1$ with $\theta_0 = \theta_{-1} = 1$, and smoothing parameter $\mu > 1$, the following holds for the output of the functions f_W^k defined in* (2.6) *-* (2.7)*:*

$$\begin{aligned}\|f_W^k(Y)\|_F \leq 2\mu\|Y\|_F \Bigg[\sum_{i=0}^{k-1} \Bigg(\Big(&\|A\|_{2\to2}(c_{1,i-1}\Lambda + c_{2,i-1}\|A\|_{2\to2}) + c_{2,i-1} \Big) \\ &\cdot \prod_{j=i}^{k-1} \Gamma_j \Bigg) + \|A\|_{2\to2}(c_{1,k-1}\Lambda + c_{2,k-1}\|A\|_{2\to2}) + c_{2,k-1} \Bigg],\end{aligned} \tag{2.17}$$

where $\{c_{1,k}\}_{k\geq0}, \{c_{2,k}\}_{k\geq0} \leq 1$ with $c_{1,-1} = c_{2,-1} = 0$, and

$$\Gamma_k = 2\big[c_{1,k}\Lambda^2 + c_{2,k}\|A\|_{2\to2}^2 + 2\|A\|_{2\to2}\Lambda(c_{1,k} + c_{2,k})\big] + 1, \tag{2.18}$$

Moreover, if $c_{1,k}\Lambda \leq 1$, $c_{1,k}\Lambda^2 \leq 1$, $c_{2,k}\|A\|_{2\to2}^2 \leq 1$, then we have the simplified upper bound

$$\|f_W^k(Y)\|_F \leq 2\mu\|Y\|_F(\|A\|_{2\to2} + 1)(\zeta_k + 1), \tag{2.19}$$

where $\zeta_k = \frac{\gamma^k - 1}{\gamma - 1}$, with $\gamma = 4(\Lambda + \|A\|_{2\to2} + 1) + 1$. In fact, it can be proven that $\Gamma_k \leq \gamma$ for all $k \geq 0$.

2.5 Lipschitzness Results

We prove that the intermediate decoder (2.10) is Lipschitz continuous w.r.t. W and explicitly calculate the Lipschitz constants, which depend on L.

Theorem 2.2. *Let f_W^L defined as in (2.10), $L \geq 2$, dictionary $W \in \mathcal{B}_\Lambda$, step sizes $0 < \{t_k^1\}_{k\geq 0}, \{t_k^2\}_{k\geq 0} \leq 1$ with $t_0^1 = t_0^2 = 1$, $t_{-1}^1 = t_{-1}^2 = 0$, step size multiplier $0 < \{\theta_k\}_{k\geq 0} \leq 1$ with $\theta_0 = \theta_{-1} = 1$, and smoothing parameter $\mu > 1$. Then, for any $W_1, W_2 \in \mathcal{B}_\Lambda$, we have*

$$\|f_{W_1}^L(Y) - f_{W_2}^L(Y)\|_F \leq K_L \|W_1 - W_2\|_{2\to 2}, \tag{2.20}$$

where

$$\begin{aligned} K_L = 2\mu\|Y\|_F \Bigg(\mu^{-1}\|A\|_{2\to 2} + \sum_{k=2}^{L} \Bigg(\Big(\max_{0\leq l\leq L-1} \Gamma_l \Big)^{L-k} \\ \cdot \sum_{i=0}^{k-2} 2 \Bigg(\Big(\|A\|_{2\to 2}(c_{1,i-1}\Lambda + c_{2,i-1}\|A\|_{2\to 2}) + c_{2,i-1} \Big) \\ \prod_{j=i}^{k-2} \Gamma_j \Bigg) + 2 \Big(\|A\|_{2\to 2}(c_{1,k-2}\Lambda + c_{2,k-2}\|A\|_{2\to 2}) + c_{2,k-2} \Big) \\ \cdot (2\Lambda c_{1,k-1} + \|A\|_{2\to 2}(c_{1,k-1} + c_{2,k-1})) + c_{1,k-1}\|A\|_{2\to 2} \Bigg) \Bigg), \end{aligned} \tag{2.21}$$

with $\{\Gamma_k\}_{k\geq 0}$, $\{c_{1,k}\}_{k\geq 0}$, $\{c_{2,k}\}_{k\geq 0}$ defined as in Lemma 2.1 and $c_{1,-1} = c_{2,-1} = 0$. Moreover, if $c_{1,k}\Lambda \leq 1$, $c_{1,k}\Lambda^2 \leq 1$, $c_{2,k}\|A\|_{2\to 2}^2 \leq 1$, for all $k \geq 0$, then we have the simplified upper bound

$$K_L \leq 2\mu\|Y\|_F \left(\|A\|_{2\to 2}(L - 1 + \mu^{-1}) + 2(\|A\|_{2\to 2} + 1)(\|A\|_{2\to 2} + 3)\kappa_L \right), \tag{2.22}$$

where

$$\kappa_L = \gamma^L \left(\frac{L-1}{\gamma(\gamma-1)} + \frac{\gamma(\gamma-2)}{(\gamma-1)^2} \right) - \frac{\gamma^2(\gamma-2)}{(\gamma-1)^2}, \tag{2.23}$$

with γ as in Lemma 2.1.

We can now prove the Lipschitzness of the main decoder defined in (2.11).

Corollary 2.3. *Let $h \in \mathcal{H}^L$ defined as in (2.12) with $L \geq 2$ and dictionary $W \in \mathcal{B}_\Lambda$. Then, for any $W_1, W_2 \in \mathcal{B}_\Lambda$, we have:*

$$\|\psi(\phi(f_{W_2}^L(Y))) - \psi(\phi(f_{W_1}^L(Y)))\|_F \leq \mu^{-1}(\Lambda + \|A\|_{2\to 2})K_L\|W_2 - W_1\|_F, \tag{2.24}$$

with K_L as in Theorem 2.2.

2.6 Covering numbers and Dudley's inequality

For a fixed number of layers $L \in \mathbb{N}$, we define the set $\mathcal{M} \subset \mathbb{R}^{n\times s}$ corresponding to the hypothesis class $\mathcal{H}^L$ to be

$$\begin{aligned} \mathcal{M} := &\{(h(y_1)|h(y_2)|\dots|h(y_s)) \in \mathbb{R}^{n\times s} : \ h \in \mathcal{H}^L\} \\ = &\{\psi(\phi((f_W^L(Y))) \in \mathbb{R}^{n\times s} : \ W \in \mathcal{B}_\Lambda\}. \end{aligned} \tag{2.25}$$

The column elements of each matrix in $\mathcal{M}$ are the reconstructions given by a decoder $h \in \mathcal{H}^L$ when applied to the measurements y_i. Since $\mathcal{M}$ is parameterized by W like $\mathcal{H}^L$ is, we have

$$\mathcal{R}_{\mathcal{S}}(\mathcal{H}^L) = \mathcal{R}_{\mathcal{S}}(\mathcal{M}) = \mathbb{E} \sup_{M \in \mathcal{M}} \frac{1}{s} \sum_{i=1}^{s} \sum_{k=1}^{n} \epsilon_{ik} M_{ik}. \tag{2.26}$$

We employ Dudley's inequality [1] in order to upper bound the right hand side of (2.26) in terms of the covering numbers of $\mathcal{M}$. Therefore,

$$\mathcal{R}_{\mathcal{S}}(\mathcal{H}^L) \leq \frac{16(\mathrm{B_{in}} + \mathrm{B_{out}})}{s} \int_0^{\frac{\sqrt{s}B_{\text{out}}}{2}} \sqrt{\log \mathcal{N}(\mathcal{M}, \|\cdot\|_F, \varepsilon)} d\varepsilon. \tag{2.27}$$

We estimate $\mathcal{N}(\mathcal{M}, \|\cdot\|_F, \varepsilon)$ by means of

Proposition 2.4. *The following estimate holds for the covering numbers of $\mathcal{M}$:*

$$\mathcal{N}(\mathcal{M}, \|\cdot\|_F, \varepsilon) \leq \left(1 + \frac{2\Lambda(\Lambda + \|A\|_{2\to 2})K_L}{\mu\varepsilon}\right)^{Nn}. \tag{2.28}$$

2.7 Generalization Error Bounds

We are now in position to deliver generalization error bounds for DECONET.

Theorem 2.5. *Let $\mathcal{H}^L$ be the hypothesis class defined in* (2.12)*. Then, for $\delta \in (0,1)$, with probability at least $1-\delta$, for all $h \in \mathcal{H}^L$, the generalization error is bounded as*

$$\begin{aligned}\mathcal{L}(h) \leq \hat{\mathcal{L}}(h) + 8(B_{\text{in}} + B_{\text{out}})B_{\text{out}}\sqrt{\frac{Nn}{s}}\sqrt{\log\left(e\left(1 + \frac{4\Lambda(\Lambda + \|A\|_{2\to 2})K_L}{\sqrt{s}B_{\text{out}}\mu}\right)\right)} \\ + 4(B_{\text{in}} + B_{\text{out}})^2\sqrt{\frac{2\log(4/\delta)}{s}},\end{aligned} \tag{2.29}$$

with K_L defined in (2.21).

Under some additional, simplifying assuptions, similar to the ones presented in Section 2.4.1, we obtain

Corollary 2.6. *Let $\mathcal{H}^L$ be the hypothesis class defined in* (2.12) *and assume that $c_{1,k}\Lambda \leq 1$, $c_{1,k}\Lambda^2 \leq 1$, $c_{2,k}\|A\|_{2\to 2}^2 \leq 1$, for all $k \geq 0$, with $\{c_{1,k}\}$, $\{c_{1,k}\} \leq 1$ defined as in Lemma 2.1. Then, for $\delta \in (0,1)$, with probability at least $1-\delta$, for all $h \in \mathcal{H}^L$, the generalization error is bounded as*

$$\begin{aligned}\mathcal{L}(h) \leq& \hat{\mathcal{L}}(h) + 8(B_{\text{in}} + B_{\text{out}})\left(B_{\text{out}}\sqrt{\frac{Nn}{s}}\sqrt{\log\left(e\left(1 + \frac{\|Y\|_F(p + qL + r\kappa_L)}{\sqrt{s}B_{\text{out}}}\right)\right)}\right. \\ &\left. + \sqrt{\frac{2\log(4/\delta)}{s}}\right),\end{aligned} \tag{2.30}$$

with κ_L as in Theorem 2.2 and p, q, $r > 0$ constants depending on $\|A\|_{2\to 2}, \Lambda, \mu$.

All the previous results are summarized in

Theorem 2.7. *Let $\mathcal{H}^L$ be the hypothesis class defined in* (2.12)*. Assume there exist pair-samples $\{(x_i, y_i)\}_{i=1}^s$, with $y_i = Ax_i + e$, $\|e\|_2 \leq \varepsilon$, for some $\varepsilon > 0$, that are drawn i.i.d. according to an unknown distribution $\mathcal{D}$, and that it holds $\|y_i\|_2 \leq \mathrm{B_{in}}$ almost surely with $\mathrm{B_{in}} = \mathrm{B_{out}}$ in the definition of the truncating function ψ. Let us further assume that for step sizes $0 < \{t_k^1\}_{k\geq 0}$, $\{t_k^2\}_{k\geq 0} \leq 1$, step size multiplier $0 < \{\theta_k\}_{k\geq 0} \leq 1$ and smoothing parameter $\mu > 1$, we have $\mu^{-1}\theta_k^{-1}t_k^1\Lambda \leq 1$, $\mu^{-1}\theta_k^{-1}t_k^1\Lambda^2 \leq 1$, $\mu^{-1}\theta_k^{-1}t_k^2\|A\|_{2\to 2} \leq 1$, for all $k \geq 0$. Then, for $\delta \in (0,1)$, with probability at least $1-\delta$, for all $h \in \mathcal{H}^L$, the generalization error is bounded as*

$$\begin{aligned}\mathcal{L}(h) \leq & \hat{\mathcal{L}}(h) + 16B_{\text{out}}^2\sqrt{\frac{Nn}{s}}\sqrt{\log\left(e\left(1+\frac{p+qL+r\kappa_L}{\sqrt{s}B_{\text{out}}}\right)\right)} \\ & + 16B_{\text{out}}\sqrt{\frac{2\log(4/\delta)}{s}},\end{aligned} \tag{2.31}$$

with κ_L as in Theorem 2.2 and constants p, q, $r > 0$ as in Corollary 2.6.

Corollary 2.8 (Informal)**.** *According to* (2.23)*, we have that L enters at most exponentially in the definition of κ_L. If we consider the dependence of the generalization error bound* (2.31) *only on L, N, s and treat all other terms as constants, we roughly have*

$$|\mathcal{L}(h) - \hat{\mathcal{L}}(h)| \lesssim \sqrt{\frac{NL}{s}}. \tag{2.32}$$

2.8 Numerical Experiments

We train and test DECONET on real-world image datasets (MNIST and CIFAR10). We consider the vectorized version of images. We report the *test MSE* (which approximates the true loss (2.13))

$$\hat{\mathcal{L}}_{test}(h) = \frac{1}{p}\sum_{i=1}^{p}\|h(\tilde{y}_i) - \tilde{x}_i\|_2^2, \tag{2.33}$$

where $\mathcal{P} = \{(\tilde{y}_i, \tilde{x}_i)\}_{i=1}^p$ is a set of p test data, not used in the training phase. We also report the *empirical generalization error* (EGE) defined by

$$\mathcal{L}_{gen} = |\hat{\mathcal{L}}_{test}(h) - \hat{\mathcal{L}}_{train}(h)|. \tag{2.34}$$

We use (2.34) – which can be explicitly computed – to evaluate GE.

2.8.1 Discussion

As illustrated in Fig. 2, the test MSEs drop as L and N increase. This is a reasonable behaviour for a standard analysis CS scenario, since the performance and reconstruction quality of (1.1) typically benefit from the (high) redundancy of the involved sparsifier and the increasing number of iterations/layers. Moreover, EGEs appear to grow like $\sqrt{NL}$, as L and N increase, which confirms our theoretical findings in Section 2.7.

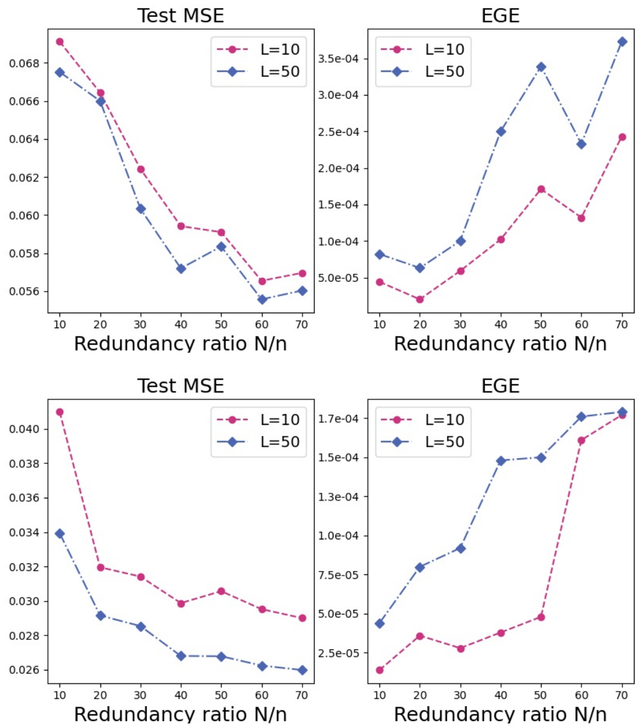


Figure 2: Performance plot for 10- and 50-layer DECONET with $m = n/4$, tested on MNIST (top) and CIFAR10 (bottom) datasets.

2.8.2 Conclusion & Future Work

In this paper, we derived a new unfolding network coined DECONET with $L \in \mathbb{N}$ layers. DECONET jointly learns a decoder for CS and a redundant sparsifying analysis operator $W \in \mathbb{R}^{N \times n}$. We estimated DECONET's generalization error (which roughly scales like $\sqrt{NL}$), employing chaining. To our knowledge, we are the first to present generalization error bounds of an unfolding network solving analysis CS. Finally, we presented numerical experiments, which confirmed our theory. As a future direction, we would like to characterize (e.g. in terms of structure) the redundant sparsifying transform that DECONET learns.

References

[1] S. Foucart and H. Rauhut. "An invitation to compressive sensing". In: *A mathematical introduction to compressive sensing*, Springer, 2013, pp 1–39.

[2] M. Elad. "Sparse and redundant representations. From theory to applications in signal and image processing". With a foreword by Alfred M. Bruckstein. Springer, New York, 2010. xx+376 pp.

[3] M. Genzel, G. Kutyniok and M. Marz. "ℓ_1-analysis minimization and generalized (co-)sparsity: when does recovery succeed?". *Appl. Comput. Harmon. Anal.* 52 (2021), pp. 82–140.

[4] M. Kabanava and H. Rauhut. "Analysis ℓ_1-recovery with frames and Gaussian measurements". *Acta Appl. Math.* 140.1 (2015), pp. 173–195.

[5] E. Candès, Y. C. Eldar, D. Needell and P. Randall. "Compressed sensing with coherent and redundant dictionaries". *Appl. Comput. Harmon. Anal.* 31.1 (2011), pp. 59–73.

[6] V. Kouni and H. Rauhut. "Spark Deficient Gabor Frame Provides A Novel Analysis Operator For Compressed Sensing". In: *Neural Information Processing*. Ed. by T. Mantoro et al. Cham: Springer International Publishing, 2021, pp. 700–708.

[7] P. G. Casazza, G. Kutyniok and F. Philipp. "Introduction to finite frame theory". *Finite frames*, 2013, pp. 1–53.

[8] R.-D. Malikiosis. "Spark deficient Gabor frames". *Pacific J. Math.* 294.1 (2018), pp. 159–180.

[9] H. B. Dang, K. Blanchfield, I. Bengtssn and D. M. Appleby. "Linear dependencies in Weyl-Heisenberg orbits". *Quantum Inf. Process.* 12.11 (2013), pp. 3449–3475.

[10] M. Vishal, Y. Li and Y. C. Eldar. "Algorithm Unrolling: Interpretable, Efficient Deep Learning for Signal and Image Processing". *IEEE Signal Process. Mag.* 38.2 (2021), pp. 18–44.

[11] A. Behboodi, H. Rauhut and E. Schnoor. "Compressive Sensing and Neural Networks from a Statistical Learning Perspective". *arXiv:2010.15658* (2020).

[12] Y. Yang, J. Sun, H. Li and Z. Xu. "ADMM-Net: A Deep Learning Approach for Compressive Sensing MRI". *arXiv:1705.06869* (2017).

[13] V. Kouni, G. Paraskevopoulos, H. Rauhut and G. C. Alexandropoulos. "ADMM-DAD Net: A Deep Unfolding Network for Analysis Compressed Sensing". In: *Int. Conf. Acoust., Speech and Signal Process.* IEEE, 2022, pp. 1506–1510.

[14] S. Shalev-Shwartz and S. Ben-David. *Understanding machine learning: From theory to algorithms.* Cambridge University Press, 2014.

[15] S. R. Becker, E. J. Candés and M. C. Grant. "Templates for convex cone problems with applications to sparse signal recovery". *Math. Program. Comput.* 3.3 (2011), pp. 165–218.

[16] C. Ma et al. "Rademacher complexity and the generalization error of residual networks". *Commun. Math. Sci.* 18.6 (2020), pp. 1755–1774.

[17] V. Kouni and Y. Panagakis. "DECONET: an Unfolding Network for Analysis-based Compressed Sensing with Generalization Error Estimates". *arXiv:2205.07050* (2022).

Vicky Kouni: Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, Greece, Panepistimioupolis 157-84, Athens, Greece.

E-mail: vicky-kouni@di.uoa.gr

Convergence rates to steady states for a heat conduction model - the chain of oscillators

Angeliki Menegaki

In memory of Dimitris Gatzouras

Abstract

In this summary we present the main results from the works [Men20, BM22] where we study the long-time behaviour of an out-of-equilibrium heat conduction model.

1 Introduction

The main objective is to find estimates on the speed of the convergence to a stationary state for a heat conducting system.

1.1 Motivation

The motivation for this study is the rigorous mathematical understanding of Fourier's law. Fourier's law is a physical macroscopic law that relates the local thermal flux $J(t,x)$ to small variations of temperature $\nabla T(t,x)$ through a proportionality constant $\kappa(T)$ known as *thermal conductivity*:

$$J(t,x) = -\kappa(T)\nabla T(t,x). \tag{1.1}$$

At the microscopic scale, matter is made out of particles assumed to evolve according to the classical laws of mechanics, and one of the goals of statistical physics is to model heat conductivity through a system of interacting atoms and to achieve a rigorous derivation of constitutive laws such as Fourier's law [BLRB00, FB19, Lep16, Dha08]. Understanding macroscopic laws of matter when starting from a microscopic system of interacting atoms is a challenge addressed to mathematicians by Hilbert in his 6^{th} problem [Hil02].

A paradigmatic set up where Fourier's law is observed to hold with high precision is when one considers a fluid in a cylindrical slab of height h and uniform cross sectional area A, coupled at the two boundaries, the top and the bottom of the cylinder, to two heat reservoirs at different temperatures. This is known as the Benard experiment [BLRB00]. The two heat reservoirs keep the system out of equilibrium and produce a stationary heat flow. If there is a non-equilibrium steady state (NESS) that is described by a phase-space measure, one would like to prove that the following limit exists:

$$0 < \kappa(N) := \lim_{N\to\infty} \frac{\langle J^N(t,x)\rangle}{(A(\Delta T/N))} < \infty \tag{1.2}$$

where N is the microscopic length of the cylinder,

$$\frac{\Delta T}{N} = \frac{T_2 - T_1}{N}$$

is the effective temperature gradient, $\langle J^N(t,x)\rangle$ is the expectation of the heat flux with respect to the non equilibrium steady state and where we write $J^N(t,x)$ to stress the dependence of J on N. The above limit allows us to define the thermal conductivity and the very existence of the limit is a formulation of Fourier's law.

Our main objective is therefore to investigate how certain quantities, such as the relaxation rates to the NESS of such systems (the spectral gap of the associated dynamics), scale with the system size, since these are crucial to making sure that the thermal conductivity has a thermodynamic limit.

1.2 Preliminaries

We give here some definitions in order to state later our main results.

Definition (Entropy and Log-Sobolev Inequalities)**.** For a probability measure μ on some Borel set Ω the entropy $\mathrm{Ent}_\mu(F)$ of a positive measurable function $F : \Omega \to \mathbb{R}_{\geq 0}$ with

$$\int_\Omega F(x) \log^+(F(x))\, d\mu(x) < \infty$$

is defined as

$$\mathrm{Ent}_\mu(F) := \int_\Omega F(x) \log\left(F(x) \Big/ \int_\Omega F(y)\, d\mu(y) \right)\, d\mu(x). \tag{1.3}$$

We say that μ satisfies a logarithmic Sobolev inequality *LSI(k)* iff

$$\mathrm{Ent}_\mu(f^2) \leq \tfrac{2}{k} \left\| \nabla f \right\|^2_{L^2(d\mu)}$$

for all smooth functions f. The *LSI(k)* implies [Led99, Prop. 2.1] that μ satisfies a spectral gap inequality (Poincaré Inequality) *SGI(k)*

$$\mathrm{Var}_\mu(f) \leq \tfrac{1}{k} \left\| \nabla f \right\|^2_{L^2(d\mu)}.$$

Logarithmic Sobolev (and other functional) inequalities are very effective tools to study the concentration of measure phenomenon and to quantify the relaxation rates, i.e. the mixing properties, of the dynamics of many-particle systems [Gro93, BE85, Led99, Led01, GZ03, ABC+00]. This is since the spectral gap (the speed of relaxation) is known to be determined by the constant in the Log-Sobolev inequalities. We define the spectral gap to be the size of the gap between 0 and the rest of the spectrum of the associated generator L which can be also characterized by

$$\lambda_S := \inf_{\substack{f\in C^\infty \\ \|\nabla f\|_{L^2(d\mu)}\neq 0}} \frac{-\langle Lf, f\rangle_{L^2(d\mu)}}{\mathrm{Var}_\mu(f)}$$

where Var_μ is the variance relative to the equilibrium measure μ.

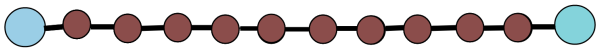


Figure 1: Homogeneous chain: Spectral gap $\sim N^{-3}$.

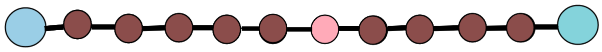


Figure 2: Chain with impurity: Spectral gap $\sim e^{-cN}$.

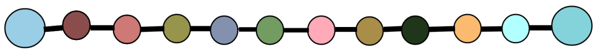


Figure 3: Disordered chain: Spectral gap $\sim e^{-cN}$.

Figure 4: The one-dimensional harmonic chain of oscillators connected to heat baths (big discs) and with various pinning potentials (differently colored discs indicate different pinning strengths).

2 Description of the model and state of the art

The model we focus on is a prototypical example of out-of-equilibrium systems and is a generalized version of the historical Fermi-Pasta-Ulam (FPU) chain. It consists of a chain of N interacting oscillators on the phase space $\mathbb{R}^{2dN}$, where the variables are q_i, p_i for $i = 1, \dots, N$: the displacements of the particles from their equilibrium positions and their momenta, respectively. Each particle has its own *pinning potential* and it interacts with its nearest neighbours through an *interaction potential*. We call H the Hamiltonian energy.

The dynamics of this model is such that the particles at the boundary are coupled to heat baths, modelled by *Langevin* (Ornstein-Uhlenbeck) processes at (possibly) different temperatures $\beta_i^{-1}, i \in \mathcal{F}$ and they are subject to friction. $\mathcal{F} \subset \{1, \dots, N\}$ here is the subset of the particles on which we impose friction and noise and we also denote by $\gamma_i > 0$ the friction strength at the i-th particle.

The time evolution is then for particles $i \in \{1, \dots, N\}$ described by a coupled system of SDEs:

$$\begin{aligned} dq_i(t) &= (\nabla_{p_i} H)dt \text{ and} \\ dp_i(t) &= \big(-\nabla_{q_i} H - \gamma_i p_i \delta_{i\in\mathcal{F}} \big)\, dt + \delta_{i\in\mathcal{F}} \sqrt{\frac{2\gamma_i}{\beta_i}}\, dW_i \end{aligned} \tag{2.1}$$

where $\gamma_i, i \in \mathcal{F}$ are the friction coefficients.

The generator of the dynamics, restricting for simplification to the 1-dimensional case and when $\mathcal{F} = \{1, N\}$ is:

$$\mathcal{L} = \sum_{i=1}^{N} (p_i \partial_{q_i} - \partial_{q_i} H \partial_{p_i}) - \gamma_1 p_1 \partial_{p_1} - \gamma_N p_N \partial_{p_N} + \gamma_1 \beta_1^{-1} \partial_{p_1}^2 + \gamma_N \beta_N^{-1} \partial_{p_N}^2.$$

Note that this operator is neither elliptic nor coercive, facts that make all the classical tools fail when it comes to the study of the regularity or the long-time behaviour of solutions.

2.1 In the literature

The non-equilibrium steady state for the purely harmonic chain, i.e. when both potentials are quadratic (harmonic), was made precise in [RLL67]. Anharmonic chains were studied in various works [JP98, EPRB99a, EPRB99b, Car07, RBT02, CEHRB18], where existence, uniqueness of a non-equilibrium steady state and exponential convergence towards it were proven in certain cases. More specifically the existence, uniqueness of a steady state and exponential convergence, hold under the assumptions that both the interaction and pinning potentials behave as polynomials near infinity and that the interaction is stronger than the pinning potential. The last assumption is important as there are some works which exhibit cases where the relaxation rate is not exponential, i.e. where there is lack of spectral gap [Hai09, HM09]. The existing results are however not quantitative, i.e. they do not give information on the scaling of these rates in terms of N, since compactness arguments are employed. Quantitative results for the spectral gap are therefore missing and even in the simplest case of the linear (harmonic) chain, the dependence on the dimension of the spectral gap was not known. Attempts have been made through hypocoercive techniques to get N-dependent estimates under certain conditions on the potentials: see the discussion in [Vil09, Section 9.2] where this question was first raised. The techniques discussed in Villani's monograph however only yield non-optimal estimates.

3 Main results

3.1 On the long time behaviour

Regarding the long time behaviour of the system, we provide explicit rates of convergence to the non-equilibrium steady state (with optimal lower bound) in a 1-dimensional weakly anharmonic scenario, i.e. when both potentials are N-dependent perturbations of the harmonic ones.

The first statement concerns a contraction property in Wasserstein-2 distance. We recall the definition of the Kantorovich-Rubinstein-Wasserstein L^2-distance $W_2(\mu, \nu)$ between two probability measures μ, ν:

$$W_2(\mu, \nu)^2 = \inf \int_{\mathbb{R}^N \times \mathbb{R}^N} |x - y|^2 d\pi(x, y)$$

where the infimum is taken over the set of all the couplings, i.e. the joint measures π on $\mathbb{R}^N \times \mathbb{R}^N$ with left and right marginals μ and ν respectively.

Theorem 3.1. *We consider a* 1*-dimensional chain of coupled oscillators with rigidly fixed edges so that the dynamics are described by the system* (2.1) *with*

$$H(p,q)=\sum_{i=1}^{N}\left(\frac{p_i^2}{2}+a\frac{q_i^2}{2}+U_{pin}^N(q_i)\right)+\sum_{i=1}^{N-1}\left(c\frac{(q_{i+1}-q_i)^2}{2}+U_{int}^N(q_{i+1}-q_i)\right)+ \\ +c\frac{q_1^2}{2}+c\frac{q_N^2}{2} \tag{3.1}$$

for $a\geq 0, c>0$ *and under the assumption that*

$$\sup_{q_i}\|\nabla^2 U_{pin}^N(q_i)\|_2,\quad \sup_{r_i}\|\nabla^2 U_{int}^N(r_i)\|_2\leq C^N \tag{3.2}$$

where $r_i=q_{i+1}-q_i$ *and* $C^N\lesssim N^{-9/2}$. [1] *For a fixed number of particles* N*, there is a unique stationary state, in particular, for initial data* f_0^1, f_0^2 *we have:*

$$W_2(P_t^*f_0^1,P_t^*f_0^2)\leq C_{a,c}N^{\frac{3}{2}}e^{-\frac{\lambda_0}{N^3}t}\,W_2(f_0^1,f_0^2) \tag{3.3}$$

for $C_{a,c},\lambda_0$ *dimensionless constants.*

The proof relies on

- an application of a generalized version of the Γ_2-calculus of Bakry-Emery [BE85] for elliptic operators recently generalized by Baudoin for hypoelliptic operators [Bau17] and
- a careful analysis of a high-dimensional matrix equation.

The generalised Γ_2-calculus allows us to prove the validity of a Log-Sobolev inequality for the invariant measure, with constant $C_N\lesssim N^3$. With this inequality in hand we also give a convergence to the stationary measure in relative entropy as in [Vil09, Section 6]. We first recall the definitions of the following functionals:

For two probability measures μ and ν on $\mathbb{R}^{2N}$ with $\nu\ll\mu$, we define the Boltzmann H functional (relative entropy)

$$H_\mu(\nu)=\int_{\mathbb{R}^{2N}}h\log h\,d\mu,\ \nu=h\mu \tag{3.4}$$

and the relative Fisher information

$$I_\mu(\nu)=\int_{\mathbb{R}^{2N}}\frac{|\nabla h|^2}{h}d\mu,\ \nu=h\mu. \tag{3.5}$$

Theorem 3.2. *We consider a weakly anharmonic* 1*-dimensional chain of coupled oscillators with rigidly fixed edges whose dynamics are described by the system* (2.1) *under the same assumptions as in the Theorem 3.1 above. For a fixed number of particles* N*, assuming that (i)* μ *is the invariant measure for* P_t *and (ii) that it satisfies a Log-Sobolev inequality with constant* $C_N>0$*, for all* $0<f\in L^1(\mu)$ *with*

$$\mathcal{E}(f)<\infty,\ \textit{and}\ \int f d\mu=1,$$

we have a convergence to the non-equilibrium steady state in the following sense:

$$H_\mu(P_tf\mu)+I_\mu(P_tf\mu)\leq\lambda_{a,c}N^3e^{-\lambda_0N^{-3}t}\Big(H_\mu(f\mu)+I_\mu(f\mu)\Big) \tag{3.6}$$

for dimensionless constants $\lambda_{a,c},\lambda_0$.

[1] This is what we call a weakly anharmonic chain of oscillators.

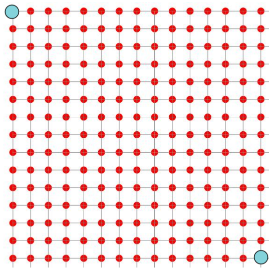


Figure 5: Spectral gap $\sim N^{-6}$.

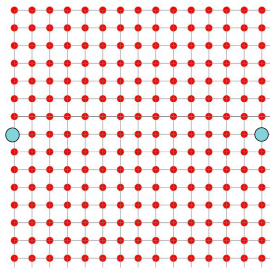


Figure 6: Spectral gap $\sim N^{-4}$.

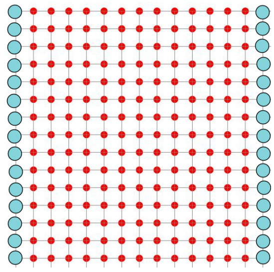


Figure 7: Spectral gap $\sim N^{-3}$.

3.2 On the spectral gap

Furthermore, we study the spectral gap for purely harmonic chains and d-dimensional grids of oscillators, and proved the optimal lower and upper bounds. We also treat non-homogeneous scenarios where the coefficients of the pinning potentials are not identical. In particular we look at chains of oscillators with an impurity (so that the particle in the middle of the chain has pinning potential significantly weaker than the pinning potential of all the other particles) as well as at disordered chains of oscillators, see Fig 1,2,3. As regards the d-dimensional grids, the spectral gap depends on which particles are exposed to friction, cf. Fig. 5,6,7. These are explained in the statement below.

Our setting is the following, we look at the system (2.1) with $\mathcal{F} \subset \{1, \dots, N\}^d$ and

$$H(q,p) = \frac{\langle p, m_{[N]^d}^{-1} p \rangle}{2} + V_{a,c}(q) \text{ where } V_{a,c}(q) = \sum_{i \in [N]^d} a_i |q_i|^2 + \sum_{i \sim j} c_{ij} |q_i - q_j|^2. \quad (3.7)$$

Theorem 3.3. *Let the positive masses m_i and interaction strengths c_i of all oscillators coincide, N^d be the number of oscillators, placed in a square grid with N oscillators on each side, and d the dimension of the network.*

- **(Homogeneous chain):** *Let the pinning strength a_i be the same for all oscillators, then*

  (i) *if two particles located at the corners $(1, \dots, 1), (N, \dots, N)$, see Fig. 5, are exposed to the same non-zero friction and non-zero diffusion, the spectral gap of the generator decays at the optimal rate N^{-3d}:*

  $$\lambda_N = \mathcal{O}(N^{-3d}).$$

  In particular for the one-dimensional chain of oscillators $\lambda_N = \mathcal{O}(N^{-3})$.

  (ii) *if the same non-zero friction and non-zero diffusion for particles located at the center of two opposite edges of the network*

  $$(1, \lceil N/2 \rceil, \dots, \lceil N/2 \rceil), (N, \lceil N/2 \rceil, \dots, \lceil N/2 \rceil),$$

  see Fig. 6, the spectral gap of the generator decays at the optimal rate $N^{-3-(d-1)}$: $\lambda_N = \mathcal{O}(N^{-3-(d-1)})$.

(iii) *if $d = 2$ and the particles exposed to the same non-zero friction are located at opposite edges of the network, the spectral gap satisfies $\lambda_N \le \mathcal{O}(N^{-5/2})$.*

- **(Chain with impurity):** *Let N be even. We assume that all masses and interaction parameters are positive and coincide and the friction parameters γ_i of the boundary particles*

  $$\partial[N]^d := \{i \in [N]^d; \exists i_n : i_n \in \{1, N\}\} \ \textit{of} \ [N]^d := \{1, .., N\}^d$$

  satisfy $\sup_{i \in \partial[N]^d} \gamma_i \in (0, c)$where c is independent of N and the friction is non-zero on at least one boundary edge. Then, if the pinning strength $a_{c_d(N)}$ at the center point $c_d(N) = (N/2, .., N/2)$ of the network is sufficiently small compared to the pinning strength of all other oscillators, the spectral gap λ_N of the generator decays at least exponentially fast in N, for all $d \ge 1$.

  In dimension 1 this rate is the optimal one.

- **(Disordered chain):** *We assume that all masses and interaction parameters are positive and coincide and the friction parameters γ_i of the particles at the boundary*

  $$\partial[\pm N]^d := \{i \in [\pm N]^d; \|i\|_\infty = N\} \ \textit{of the network} \ [\pm N]^d := \{-N, ..., N\}^d$$

  satisfy $\sup_{i \in \partial[\pm N]^d} \gamma_i \in (0, c)$ where c is independent of N and the friction is non-zero on at least one boundary edge. Then, if the pinning strengths are iid random variables distributed according to some compactly supported density $\rho \in C_c(0, \infty)$, the spectral gap λ_N of the generator decays exponentially fast in N, for all $d \ge 1$, for all but finitely many N.

The method of proof relies on a new approach for studying non-symmetric spectral problems that reduces the problem to a spectral analysis of discrete Schrödinger operators. Using a *Wigner matrix* representation we reduce the study of this high dimensional spectral analysis to the study of resolvents involving only the heat bath sites.

References

[ABC$^+$00] C. Ané, S. Blachère, D. Chafa(*i*) , P. Fougères, I. Gentil, F. Malrieu, C. Roberto, and G. Scheffer. *Sur les inégalités de Sobolev logarithmiques*, volume 10 of *Panoramas et Synthèses [Panoramas and Syntheses]*. Société Mathématique de France, Paris, 2000. With a preface by Dominique Bakry and Michel Ledoux.

[Bau17] F. Baudoin. Bakry-Émery meet Villani. *J. Funct. Anal.*, 273(7):2275–2291, 2017.

[BE85] D. Bakry and M. Émery. Diffusions hypercontractives. In *Séminaire de probabilités, XIX, 1983/84*, volume 1123 of *Lecture Notes in Math.*, pages 177–206. Springer, Berlin, 1985.

[BLRB00] F. Bonetto, J. L. Lebowitz, and L. Rey-Bellet. Fourier's law: a challenge to theorists. In *Mathematical physics 2000*, pages 128–150. Imp. Coll. Press, London, 2000.

[BM22] Simon Becker and Angeliki Menegaki. The optimal spectral gap for regular and disordered harmonic networks of oscillators. *Journal of Functional Analysis*, 282(2):109286, 2022.

[Car07] P. Carmona. Existence and uniqueness of an invariant measure for a chain of oscillators in contact with two heat baths. *Stochastic Process. Appl.*, 117(8):1076–1092, 2007.

[CEHRB18] N. Cuneo, J.-P. Eckmann, M. Hairer, and L. Rey-Bellet. Non-equilibrium steady states for networks of oscillators. *Electron. J. Probab.*, 23:28 pp., 2018.

[Dha08] A. Dhar. Heat transport in low-dimensional systems. *Advances in Physics*, 57, 08 2008.

[EPRB99a] J.-P. Eckmann, C.-A. Pillet, and L. Rey-Bellet. Entropy production in nonlinear, thermally driven Hamiltonian systems. *J. Statist. Phys.*, 95(1-2):305–331, 1999.

[EPRB99b] J.-P. Eckmann, C.-A. Pillet, and L. Rey-Bellet. Non-equilibrium statistical mechanics of anharmonic chains coupled to two heat baths at different temperatures. *Comm. Math. Phys.*, 201(3):657–697, 1999.

[FB19] P. Flandrin and C. Bernardin, editors. *Fourier and the Science of Today / Fourier et la Science d'aujourd'hui*, volume 20, Issue 5. Comptes Rendus Physique, 2019.

[Gro93] L. Gross. Logarithmic Sobolev inequalities and contractivity properties of semigroups. In *Dirichlet forms (Varenna, 1992)*, volume 1563 of *Lecture Notes in Math.*, pages 54–88. Springer, Berlin, 1993.

[GZ03] A. Guionnet and B. Zegarlinski. Lectures on logarithmic Sobolev inequalities. In *Séminaire de Probabilités, XXXVI*, volume 1801 of *Lecture Notes in Math.*, pages 1–134. Springer, Berlin, 2003.

[Hai09] M. Hairer. How hot can a heat bath get? *Comm. Math. Phys.*, 292(1):131–177, 2009.

[Hil02] D. Hilbert. Mathematical problems. *Bull. Amer. Math. Soc.*, 8(10):437–479, 07 1902.

[HM09] M. Hairer and J. C. Mattingly. Slow energy dissipation in anharmonic oscillator chains. *Comm. Pure Appl. Math.*, 62(8):999–1032, 2009.

[JP98] V. Jakšić and C.-A. Pillet. Ergodic properties of classical dissipative systems i. *Acta Math.*, 181(2):245–282, 1998.

[Led99] M. Ledoux. Concentration of measure and logarithmic Sobolev inequalities. In *Séminaire de Probabilités, XXXIII*, volume 1709 of *Lecture Notes in Math.*, pages 120–216. Springer, Berlin, 1999.

[Led01] M. Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001.

[Lep16] S. Lepri. *Thermal Transport in Low Dimensions: From Statistical Physics to Nanoscale Heat Transfer*, volume 921. 01 2016.

[Men20] A. Menegaki. Quantitative Rates of Convergence to Non-equilibrium Steady State for a Weakly Anharmonic Chain of Oscillators. *J. Stat. Phys.*, 181(1):53–94, 2020.

[RBT02] L. Rey-Bellet and L. E. Thomas. Exponential convergence to non-equilibrium stationary states in classical statistical mechanics. *Comm. Math. Phys.*, 225(2):305–329, 2002.

[RLL67] Z. Rieder, J. L. Lebowitz, and E. Lieb. Properties of a harmonic crystal in a stationary nonequilibrium state. *Journal of Mathematical Physics*, 8(5):1073–1078, 1967.

[Vil09] C. Villani. Hypocoercivity. *Mem. Amer. Math. Soc.*, 202(950):iv+141, 2009.

Angeliki Menegaki: Institut des Hautes Études Scientifiques, 35 Rte de Chartres, 91440, Bures-sur-Yvette, France.

E-mail: menegaki@ihes.fr

A generic result on the Hardy space H^1

Vassili Nestoridis

In memory of Dimitri Gatzouras

Abstract

According to the Hardy's inequality, if f is a holomorphic function on the unit disc of class H^1, then the sequence $a(f)$ of the Taylor coefficients of the primitive of f belongs to the space ℓ^1. We show that generically for all f in H^1 the sequence $a(f)$ is outside any ℓ^p space smaller than ℓ^1 i.e. with $0 < p < 1$; thus, $a(f) \in \ell^1 \setminus \Big(\bigcup\limits_{0<p<1} \ell^p \Big)$ holds generically for every f in H^1.

1 Introduction

We start with a question asked by Dimitri Gatzouras about extendability of analytic curves in the plane. Next we use Baire's theorem to prove our main result, that generically for every function f in the Hardy class H^1 in the open unit disc D of the complex plane $\mathbb{C}$, the sequence $a(f)$ of the Taylor coefficients of the primitive of f lies in $\ell^1 \setminus \Big(\bigcup\limits_{0<p<1} \ell^p \Big)$.

Baire's theorem was also used to prove generic existence of universal Taylor series in D or more generally in any simply connected domain Ω in $\mathbb{C}$ ([9], [10], [6]). The proof except Baire's theorem, uses Mergelyan's theorem that, for every compact set $K \subset \mathbb{C}$ with connected complement, every function f in $A(K)$, (that is, continuous on K and holomorphic in K^0), can be uniformly on K approximated by polynomials. There is no satisfactory version of Mergelyan's theorem is several complex variables. That is why the theory of universal Taylor series in several variables is less developed.

In order to prove generic existence of universal Taylor series on a product $\Omega = \Omega_1 \times \cdots \times \Omega_n$ of planar simply connected domains $\Omega_1, \ldots, \Omega_n$, except Baire's theorem we need a Mergelyan's type theorem, asserting that, if the planar compact sets $K_1, \ldots, K_n$ have connected complements, then every function f in $A(K_1 \times \cdots \times K_n)$ can be uniformly approximated by polynomials. Such a statement was claimed to be true in [3]; however, recently a counterexample was found in [2]. This led to the definition of a new function algebra $A_D(K)$ for $K \subset \mathbb{C}^n$ compact. This algebra $A_D(K)$ contains all uniform on K limits of polynomials and is contained in $A(K)$. If $K_1, \ldots, K_n$ are planar compact sets with connected complements, then every function f in $A_D(K_1 \times \cdots \times K_n)$ can, indeed, be uniformly approximated by polynomials. The algebra $A_D(K)$ consists of all functions continuous on K and holomorphic in every analytic disc in K. Thus, replacing $A(K)$ by $A_D(K)$, we can establish generic existence of universal Taylor series on products $\Omega = \Omega_1 \times \cdots \times \Omega_n$ of planar simply connected domains $\Omega_1, \ldots, \Omega_n$. The universal approximation holds on products $K = K_1, \times \cdots \times K_n$ of planar compact sets $K_1, \ldots, K_n$ with connected complements, such that $K \cap \Omega = \emptyset$ and it is realized by the partial sums of the Taylor development of the universal function according to any a-priori fixed enumeration of the monomials in the Taylor expansion [4], [5].

2 A question of D. Gatzouras

A curve $\gamma : I \to \mathbb{R}^2$, where I is an open interval, is called analytic if $\gamma'(t_0) \neq 0$ and $\gamma(t) = \sum\limits_{n=0}^{\infty} a_n \cdot (t - t_0)^n$ for some $a_n = a_n(t_0) \in \mathbb{R}^2$ holds on an open interval around any point t_0 in I. Certainly among all parametrizations of any given curve there are parametrizations where the previous definition does not hold. If for a parametrization $I \ni t \to \gamma(t) \in \mathbb{R}^2$ the above definition holds, then the curve is called analytic and the parameter t is called a conformal parameter for the curve. In [11] we prove that for any analytic curve the arclength s is a global conformal parameter. So, if we have an analytic curve on an open interval in order to investigate if there is an analytic extension one could consider the parametrization by arc length s and examine if there is an extension which is analytic with respect to s. Several examples are given in [11] but initially the curves $y = x^a$, $x \in (0, +\infty)$, for $a > 0$ were not examined. D. Gatzouras suggested that this example should be included in the list of examples and he asked the question if this curve can be analytically extended beyond the point $(0, 0)$. If $a = 1, 2, \ldots$ is a natural number, then obviously the curve is continued over $(-\infty, +\infty)$ as the graph of the function $y(x) = x^a$ and x is a conformal parameter. For $a = \frac{1}{1}, \frac{1}{2}, \frac{1}{3}, \ldots$ to be the inverse of a natural number, then writing $x = y^{1/a}$ we see that it has an analytic extension for $y \in (-\infty, +\infty)$ and y is a conformal parameter. For the remaining values of $a > 0$, my feeling was that the curve is not analytically extended beyond $(0, 0)$ but I did not see how to prove it; that is why I asked for help from John Pardon. Indeed, if the curve is extendable analytically beyond the point $(0, 0)$, considering the parametrization by arc length s such that $\gamma(s = 0) = (0, 0)$ we see that, since $\left.\frac{d\gamma}{ds}\right|_{s=0} = \left(\left.\frac{dx}{ds}\right|_{s=0}, \left.\frac{dy}{ds}\right|_{s=0} \right) \neq 0$, at least one of the derivatives $\left.\frac{dx}{ds}\right|_{s=0}$ and $\left.\frac{dy}{ds}\right|_{s=0}$ is non zero.

Assume that $\left.\frac{dx}{ds}\right|_{s=0}$ is non zero. By assumption γ, x and y are expressed as power series with strictly positive radius of convergence with center $s = 0$. Thus, the function $s \to x(s)$ defined on one open interval with center $s = 0$ has a holomorphic extension on a disc in $\mathbb{C} = \mathbb{R}^2$ centered at $O = (0, 0) = (0 + i \cdot 0)$. Since the derivative of this function is not zero at $s = 0$, it follows that on a smaller disc centered at 0 this function is invertible and the inverse function is holomorphic, hence representable as a power series of x. The derivative of this function is non-zero. Since s is a power series of x in a small interval and γ is a power series of s, it follows that γ is a power series of x and the derivative $\frac{d\gamma}{dx}$ is not zero on a small interval around $x = 0$. Thus, on this small interval around $x = 0$ the function $x \to y(x) = \mathrm{Im}\gamma(s(x))$ has a holomorphic extension on a disc centered to $O = (0 + i \cdot 0) = (0, 0)$. Thus, for every natural number n the derivative $\frac{d^n y}{dx^n}$ should converge to a finite value as $x \to 0^+$. Since $y = x^a$, and $\dfrac{dy}{dx} = ax^{a-1}$, it follows easily that a is equal to a natural number; otherwise for $n = [a]+1$ the $\lim\limits_{x \to 0^+} \dfrac{d^n y}{dx^n}(x)$ is equal to ∞.

So in this case a is a natural number, which is a contradiction.

If $\left.\dfrac{dy}{ds}\right|_{s=0} \neq 0$, then a similar argument shows that a is the inverse of a natural number, which is also absurd.

Thus, we have proven the following

Proposition 2.1. *The curve $y = x^a$, $x \in (0, +\infty)$ for $a > 0$ can be analytically continued beyond the point $(0,0)$, if and only if, a is a natural number or the inverse of a natural number.*

All this has been added in [11] and in the acknowledgement the names D. Gatzouras and J. Pardon are included.

3 Main result

In this section we apply Baire's theorem and we prove our main result which is a new generic result on the Hardy space H^1 on the open unit disc D of the complex plane $\mathbb{C}$. A function f holomorphic on D belongs to H^1 if $\|f\|_{H^1} \equiv \sup\limits_{0<r<1} \dfrac{1}{2\pi}\displaystyle\int_0^{2\pi} |f(re^{i\theta})|d\theta$ is finite. The space H^1 endowed with the norm$\|\ \|_{H^1}$ is a complex Banach space. The set of polynomials is a dense subset of H^1 ([1]). According to the Hardy's inequality, if $f \in H^1$ and $a(f)$ denotes the sequence of the Taylor coefficients of the primitive $F(f)$ of f satisfying $F(f)(0) = 0$ (and $F'(f) = f$), then $a(f)$ belongs to ℓ^1.

Proposition 3.1. *Let $0 < p < 1$. Then there exists $f = f_p$ in H^1, such that $a(f) \notin \ell^p$.*

Proof. Let $f(z) = \dfrac{1}{(1-z)^\gamma}$, $\gamma > 0$. Then $f \in H^1$ if and only if $\gamma < 1$. Developing in Taylor series $f(z) = \sum\limits_{n=0}^{\infty} \delta_n z^n$ where $\delta_n = \dbinom{n+\gamma-1}{n} = \dfrac{\gamma(\gamma+1)\cdots(\gamma+n-1)}{n!} \approx \dfrac{n^{\gamma-1}}{\Gamma(\gamma)}$ [12].

Thus, $a(f) = \left(a_n\right)_{n=0}^{\infty}$ and $a_0 = 0$, $a_n = \dfrac{\delta_{n-1}}{n} \approx \dfrac{1}{n^{2-\gamma}}$ $(n > 0, n \to +\infty)$ and $|a_n|^p \approx \dfrac{1}{n^{p(2-\gamma)}}$.

For $\dfrac{1}{2} < p < 1$ choose $\gamma \in \left[\dfrac{2p-1}{p}, 1\right)$ and hence $(2-\gamma)p \leq 1$.

It follows that $a(f) \notin \ell^p$.

For $0 < p < \dfrac{1}{2}$ for any $\gamma \in (0,1)$ we have $a(f) \notin \ell^p$.

The proof is complete. ■

Definition 3.2. Let $0 < p < 1$. Denote by Λ_p the set of $f \in H^1$, such that, $a(f) \notin \ell^p$.

Proposition 3.3. *Let $0 < p < 1$. Then Λ_p is a subset of H^1 which is dense and G_δ.*

Proof. Proposition 3.1 shows that $\Lambda_p \neq \emptyset$. Let $f \in \Lambda_p$. If P is a polynomial then $f + P \in \Lambda_p$. Since the set of polynomials is dense in H^1 ([1]), it follows that the set $\{f + P : P \text{ polynomial}\}$ is dense in H^1. Since the last set is contained in Λ_p, it follows that Λ_p is dense in H^1.

In order to show that Λ_p is a G_δ is suffices to prove that $H^1 \smallsetminus \Lambda_p$ is a denumerable union of closed subsets of H^1.

For M and N natural numbers we consider the set

$$E_{M,N} = \left\{ f \in H^1 : f(z) = \sum_{n=0}^{\infty} \eta_n(f) z^n,\ \sum_{n=0}^{N} \frac{|\eta_n(f)|^p}{(n+1)^p} \leq M \right\}.$$

Then $H^1 \smallsetminus \Lambda_p = \bigcup\limits_M \Big[\bigcap\limits_N E_{M,N} \Big]$.

We verify that each set $E_{M,N}$ is a closed subset of H^1. Indeed, let $f_k \in E_{M,N}$ be a sequence converging in H^1 to some $f \in H^1$. Then f_k converges uniformly on compacta of D to f, which implies $\eta_n(f_k) \xrightarrow[k \to \infty]{} \eta_n(f)$ for every $n = 0, 1, 2, \ldots$ ([1]). Since $\sum\limits_{n=0}^{N} \frac{|\eta_n(f_k)|^p}{(n+1)^p} \leq M$ for all k, it follows $\sum\limits_{n=0}^{N} \frac{|\eta_n(f)|^p}{(n+1)^p} \leq M$; that is, $f \in E_{M,N}$ and the set $E_{M,N}$ is closed in H^1. The same holds for the intersections $\bigcap\limits_N E_{M,N}$ and their denumerable union $H^1 \smallsetminus \Lambda_p$ is an F_σ. The proof is complete. ■

Theorem 3.4. *The set* $\bigcap\limits_{0<p<1} \Lambda_p$ *is a* G_δ *and dense subset of* H^1. *It follows that for the generic function* f *in* H^1 *the sequence* $a(f)$ *belongs to* $\ell^1 \smallsetminus \Big[\bigcup\limits_{0<p<1} \ell^p \Big]$.

Proof. Applying Baire's theorem to the complete space H^1 we find that $\bigcap\limits_{n=2}^{\infty} \Lambda_{1-\frac{1}{n}}$ is a G_δ and dense subset of H^1. Since the family of ℓ^p spaces is increasing, it follows that $\bigcap\limits_{n=2}^{\infty} \Lambda_{1-\frac{1}{n}} = \bigcap\limits_{0<p<1} \Lambda_p$; this gives the first part of the statement of Theorem 3.4. This combined with the fact that $a(f) \in \ell^1$ for all $f \in H^1$ ([1]), completes the proof. ■

4 Baire's theorem, Mergelyan theorem and Universality

Baire's theorem combined with Mergelyan theorem yields generic existence of universal Taylor series on simply connected planar domains.

Definition 4.1. Let K be a compact subset of C^n, $n \geq 1$. Then $P(K)$ denotes the set of all uniform on K limits of polynomials and $A(K)$ denotes the set of functions $f : K \to \mathbb{C}$ continues on K and holomorphic in K^0.

If $K^0 = \emptyset$ then $A(K) = C(K)$. The inclusion $P(K) \subset A(K)$ always hold.

Theorem 4.2. *(Mergelyan [8]). Let* $K \subset \mathbb{C}$ *be a compact set such that* $\mathbb{C} \smallsetminus K$ *is connected. Then* $P(K) = A(K)$.

Definition 4.3. Let $\Omega \subset \mathbb{C}$ be a simply connected domain and $\zeta \in \Omega$. A function $f : \Omega \to \mathbb{C}$ holomorphic on Ω belongs to the class $\mathcal{U}(\Omega, \zeta)$ of universal Taylor series with center ζ, if for every compact set $K \subset \mathbb{C} \setminus \Omega$ with $\mathbb{C} \setminus K$ connected and every $h \in A(K)$, there is a sequence $S_{\lambda_n}(f, \zeta)(z)$ of partial sums of the Taylor development of f with center ζ which converges uniformly on K towards h.

Theorem 4.4. *([9], [10], [6]). If $\Omega \subset \mathbb{C}$ is a simply connected domain and $\zeta \in \Omega$, then the class $\mathcal{U}(\Omega, \zeta)$ is a dense and G_δ subset of the space $H(\Omega)$ of holomorphic functions on Ω endowed with the topology of uniform convergence on compacta of Ω.*

The proof of Theorem 4.4 uses Baire's theorem combined with Mergelyan's theorem. One of the uses of Mergelyan theorem in this proof is the following.

Let $K_1 \subset \Omega$ be a compact set with $\mathbb{C} \setminus K_1$ connected and let $\varphi \in H(\Omega)$. Let $K_2 \subset \mathbb{C} \setminus \Omega$ be a compact set with $\mathbb{C} \setminus K_2$ connected and let h be a polynomial. We need to find a polynomial P, such that, $\sup_{z \in K_1} |P(z) - \varphi(z)| < \epsilon$ and $\sup_{z \in K_2} |P(z) - h(z)| < \frac{1}{s}$, where $\epsilon > 0$ and $s > 0$ are given.

Indeed, since $K_1 \cap K_2 = \emptyset$ the union $K_1 \cup K_2$ has also connected complement. Thus Mergelian's theorem applies for the compact set $K = K_1 \cup K_2$. The function $H : K \to \mathbb{C}$ defined by $H|K_1 = \varphi$ and $H|K_2 = h$ belongs to $A(K)$; thus, there exists a polynomnial P such that $\sup_{z \in K} |P(z) - H(z)| < \min\left(\epsilon, \frac{1}{s}\right)$.

This yields the desired result.

As mentioned in the introduction there is no satisfactory Mergelyan's theorem in several variables and the theory of universal Taylor series in $\mathbb{C}^n$ is less developed. If we wish to obtain existence of universal Taylor series on products $\Omega = \Omega_1 \times \cdots \times \Omega_n$ of planar simply connected domains $\Omega_1, \ldots, \Omega_n$, we need a Mergelyan's type theorem of the following form.

"If $K_1, \ldots, K_n$ are planar compact sets with connected complements $\mathbb{C} \setminus K_i$, then $A(K) = P(K)$ where $K = K_1 \times \cdots \times K_n$".

Counterexample 4.5. Let $n = 2$, $K_1 = \{0\}$ and $K_2 = \overline{D}$ be the closed unit disc in $\mathbb{C}$. Then $\mathbb{C} \setminus K_1$ and $\mathbb{C} \setminus K_2$ are connected. Let $h : K_1 \times K_2 \to \mathbb{C}$ be the function $h(z_1, z_2) = |z_2|$ for all $(z_1, z_2) \in K_1 \times K_2 = \{0\} \times \overline{D}$. We have $h \in A(K_1 \times K_2) = C(K_1 \times K_2)$ because the interior of $K_1 \times K_2$ in $\mathbb{C}^2$ is empty. We show that $h \notin P(K_1 \times K_2)$; indeed, if a sequence of polynomials converges uniformly on $K_1 \times K_2$ to h, then by Wieirstrass theorem the function $|z_2| = h(0, z_2)$ should be holomorphic on the open unit disc D, which is absurd. Thus, $h \notin P(K_1 \times K_2)$ and $P(K_1 \times K_2) \neq A(K_1 \times K_2)$. ■

A careful examination of the previous counterexample leads to the following definition ([2]).

Definition 4.6. Let $K \subset \mathbb{C}^n$ be compact. A function $f : K \to \mathbb{C}$ belongs to the class $A_D(K)$, if it is continuous on K and the following holds:

For every injective holomorphic mapping $\Phi : D \to K$ on an open disc $D \subset \mathbb{C}$ the composition $f \circ \Phi$ is holomorphic on D.

If $n = 1$ then $A_D(K) = A(K)$. For $n > 1$, this is no longer true. For instance the function h in Counterexample 4.5 belongs to $A(K_1 \times K_2) \setminus A_D(K_1 \times K_2)$. In general we have the inclusion $P(K) \subset A_D(K) \subset A(K)$. Furthermore, if $\overline{O}(K)$ denotes the set of uniform on K limits of functions holomorphic in (varying) open sets containing K, then we have $P(K) \subset \overline{O}(K) \subset A_D(K) \subset A(K)$. Thus, the algebra $A_D(K)$ is better for approximation than $A(K)$. Mergelyan type theorems will be those giving conditions assuring that $A_D(K)$ coincides with $P(K)$ or $\overline{O}(K)$.

I tried to prove that if $K_1, \ldots, K_n$ are planar compact sets with connected complements, then $A_D(K) = P(K)$, where $K = K_1 \times \cdots \times K_n$, but I failed. My impression was that we would need several months in order to prove this. However, Myrto Manolaki in three days at Oberwolfach gave a proof of this ([2]).

Theorem 4.7. *Let $K_1, \ldots, K_n$ be planar compact sets with connected complements. Then $A_D(K) = P(K)$, where $K = K_1 \times \cdots K_n$.*

Theorem 4.7 allows to prove generic existence of universal Taylor series on products $\Omega = \Omega_1 \times \cdots \times \Omega_n$ of planar simply connected domain $\Omega_1, \ldots, \Omega_n$, ([4],[5]).

Theorem 4.8. *Let $\Omega_1, \ldots, \Omega_n$ be planar simply connected domains and $\zeta \in \Omega = \Omega_1 \times \cdots \times \Omega_n$. If f is holomorphic in Ω, then $S_N(f, \zeta)(z)$, $N = 0, 1, 2, \ldots$ denote the sequence of partial sums of the Taylor development of f with center ζ following any a-priori given enumeration of the monomials in this development.*

There exists a holomorphic function f on Ω, such that, for every compact planar sets $K_1, \ldots, K_n$ with connected complements, such that $K \cap \Omega = \emptyset$ where $K = K_1 \times \cdots \times K_n$, the sequence $S_N(f, \zeta)(z)$, $N = 0, 1, 2, \ldots$ is uniformly dense on $A_D(K)$. The set of such functions f is a dense and G_δ subset of the space $H(\Omega)$ of holomorphic functions on Ω endowed with the topology of uniform convergence on compacta of Ω.

When Ω is planar the functions in the class $\mathcal{U}(\Omega, \zeta)$ have very wild properties; see for example [6], [7], as well as, the works of Stephen Gardiner at al where potential theory is used to address these wild properties. It is natural to ask the question, if in several variables, the functions f of Theorem 4.8 satisfy similar wild properties. For instance the class $\mathcal{U}(\Omega, \zeta)$ is independent of the center ζ in the simply connected domain Ω. Is there an analogue in several variables in the frame of Theorem 4.8?

Acknowledgement. I would like to thank A. Siskakis for the proof of Proposition 3.1.

References

[1] Duren, P. L., Theory of H^p spaces, Academic Press, New York and London, 1970.

[2] Falco, J., Gauthier, P. M., Manolaki M., Nestoridis V., A function algebra providing new Mergelyan type theorems in several complex variables, Adv. Math. 381 (2021) paper No 107649, 31pp see also arxiv: 1901.01339.

[3] Gamelin, T. W. and Garnett, J. Constructive techniques in rational approximation, Trans. Amer. Math. Soc. (1969), 143, 187-200.

[4] Kioulafa, K., Kotrovolis, G. and Nestoridis, V., Universal Taylor series on products of planar domains, Complex Anal. Synerg. 7 (2021), no 2 paper No. 19, 7 pp see also arxiv: 1909.03521.

[5] Kotsovolis, G., Partially smooth Universal Taylor series on products of simply connected domains, Monatsh. Math. 193 (2020), no 3, 657-669, see also arxiv: 1910.01759.

[6] Melas, A. and Nestoridis, V., Universality of Taylor series as a generic property of holomorphic functions, Adv. Math. 157 (2001), 138-176.

[7] Melas, A., Nestoridis, V. and Papadoperakis, I., Growth of coefficients of universal Taylor series and comparison of two cloasses of functions, J. Anal. Math. 73 (1997), 187-202.

[8] Mergelyan, S. N. Uniform approximations to functions of a complex variable, Amer. Math. Soc. Translation (1954), no 101, 99 pp. (Uspekhi Mat. Nauk, 7:2 (1952) pp. 31-122.

[9] Nestoridis, V. Universal Taylor series, Ann. Inst. Fourier 46 (1996), 1293-1306.

[10] Nestoridis, V., An extension of the notion of universal Taylor series, Computational Methods and Function Theory CMFT' 97, Nicosia, Cyprus, 1997, 421-430.

[11] Nestoridis, V. and Papadopoulos, A., Arc length as a conformal parameter for locally analytic curves, J.M.A.A. 445 (2017), no 2, 1505-1515, see also arxiv:1508.07694.

[12] Zygmund, A., Trigonometrtic series, 2nd edition, Cambridge University Press, Cambridge, 1979.

Keywords and phrases: analytic curve, Hardy space H^1, ℓ^p sequence spaces, Baire's theorem, Mergelyan's theorem, universal Taylor series.

AMS classification numbers: primary 30H10, secondary 30B40, 30K05, 32A17, 32A38.

Vassili Nestoridis: National and Kapodistrian University of Athens, Department of Mathematics, Panepistemiopolis, 157 84, Athens, Greece.

E-mail: vnestor@math.uoa.gr

On the inversion of the Laplace transform

Nickos Papadatos

In Memory of Dimitris Gatzouras

Abstract

The Laplace transform is a useful and powerful analytic tool with applications to several areas of applied mathematics, including differential equations, probability and statistics. Similarly to the inversion of the Fourier transform, inversion formulae for the Laplace transform are of central importance; such formulae are old and well-known (Fourier-Mellin or Bromwich integral, Post-Widder inversion). The present work is motivated from an elementary statistical problem, namely, the unbiased estimation of a parametric function of the scale in the basic model of a random sample from exponential distribution. The form of the uniformly minimum variance unbiased estimator of a parametric function $h(\lambda)$, as well as its variance, are obtained as series in Laguerre polynomials and the corresponding Fourier coefficients, and a particular application of this result yields a novel inversion formula for the Laplace transform.

Key words and phrases: Exponential Distribution, Unbiased Estimation; Laplace Transform; Laguerre Polynomials.

1 Introduction

For a function $u : (0, \infty) \to \mathbb{R}$, its Laplace transform is defined by the integral

$$\phi(\lambda) = \int_0^\infty \exp(-\lambda x) u(x) \mathrm{d}x, \tag{1.1}$$

provided that there exists $\lambda_0 \geq 0$ such that

$$\int_0^\infty \exp(-\lambda_0 x)|u(x)|\mathrm{d}x < \infty.$$

There is a second version of the Laplace transform, related to probability measures μ supported in (a subset of) $[0, \infty)$, namely,

$$\phi_\mu(\lambda) = \int_{[0,\infty)} \exp(-\lambda x) \mathrm{d}\mu(x); \tag{1.2}$$

it is just a matter of notation to express $\phi_\mu(\lambda)$ as $\mathbb{E} \exp(-\lambda X)$ where the nonnegative random variable X has distribution function $F(x) = \mu([0, x])$, $x \geq 0$, and $\mathbb{E}$ denotes expectation. In this setup, $\phi_\mu(\lambda)$ is denoted as $M_X(-\lambda)$ and it is called *the moment generating function of* X. It is clear that formulae (1.1) and (1.2) coincide if X has a density u (w.r.t. to Lebesgue measure on $[0, \infty)$). An inversion formula for the probabilistic version (1.2) can be found in Billingsley (1995) or Schilling *et al* (2012), and it is based on an ingenious application of the law of large numbers. The formula can be written as ($x > 0$)

$$\mu\Big([0,x)\Big) + \frac{1}{2}\mu\Big(\{x\}\Big) = \lim_N \sum_{k=0}^{N} \frac{(-1)^k}{k!}\Big(\frac{N}{x}\Big)^k \phi_\mu^{(k)}\Big(\frac{N}{x}\Big). \tag{1.3}$$

Regarding (1.1), it is known from Lerch (1903) that the transformation $u \to \phi$ is one to one. Furthermore, there are two well-known inversion formulae for (1.1), namely, the Fourier-Mellin or Bromwich integral (see Boas (1983), Cohen (2007)),

$$u(x) = \frac{1}{2\pi i}\lim_{T\to\infty}\int_{\gamma-iT}^{\gamma+iT} \exp(sx)\phi(s)\mathrm{d}s, \tag{1.4}$$

where $\gamma \geq 0$ is greater than the real part of every pole of (the analytic extension of) ϕ, and the Post (1930) or Post-Widder formula (see Widder (1946), Post (1930), Cohen (2007)),

$$u(x) = \lim_n \frac{(-1)^n}{n!}\Big(\frac{n}{x}\Big)^{n+1}\phi^{(n)}\Big(\frac{n}{x}\Big). \tag{1.5}$$

The above inversions hold under some mild restrictions, e.g., (1.4) is satisfied for almost all $x \in (0,\infty)$ (clearly, this is the best we can expect, but the formula in itself is complicated and, so, inconvenient for purposes of computation, as can be seen when applied to trivial exemplary cases), and (1.5) holds at continuity points of u, provided that u is smooth in pieces and that the growth of $|u|$ at infinity is at most of exponential order.

The present work is motivated from an elementary statistical inference problem which, at a first glance, seems to be unrelated to Laplace inversion. The problem is to find the minimum variance unbiased estimator of a given parametric function $h(\lambda)$, based on a random sample $X_1,\ldots,X_n$ from Exp(λ), with $\lambda > 0$ unknown, or, more generally, from $\Gamma(a,\lambda)$ with $a > 0$ fixed and known and $\lambda > 0$ an unknown parameter (for the definitions see Section 2). The main result provides necessary and sufficient conditions on h so that a solution of this problem exists, and shows that the solution (whenever exists) can be presented as a series in Laguerre polynomials,

$$L_n(x) = \sum_{k=0}^{n}(-1)^k\binom{n}{k}\frac{x^k}{k!}. \tag{1.6}$$

A particular application of the main result yields a novel inversion formula for the Laplace transform; see Section 3.

2 On the best unbiased estimator of a parametric function of the scale parameter in exponential/gamma models

2.1 Preliminaries and a simple parametric inference problem

The probability density of the exponential distribution, Exp(λ), is given by

$$f_\lambda(x) = \lambda\exp(-\lambda x), \quad x > 0,$$

while the Gamma distribution, $\Gamma(a, \lambda)$, has probability density

$$f_\lambda(x) = \frac{\lambda^a}{\Gamma(a)} x^{a-1} \exp(-\lambda x), \quad x > 0, \tag{2.1}$$

where $a > 0$ and $\lambda > 0$ are positive constants, so that $\mathrm{Exp}(\lambda) \equiv \Gamma(1, \lambda)$.

From now on, we suppose that $a > 0$ is known (given), and we assume that $\lambda > 0$ is the (unique) unknown parameter to be estimated from the data. More generally, we wish to estimate an arbitrary parametric function $h(\lambda)$ by using a suitable choice of an estimator

$$T = T(X_1, \ldots, X_n),$$

where T is a real valued measurable function with domain $(0, \infty)^n$ and $X_1, \ldots, X_n$ are iid (independent, identically distributed) random variables with density (2.1). Of course, the actual value of T (when $X_1 = x_1, \ldots, X_n = x_n$) must not vary with λ, but T may depends on n or a (since both are fixed and known).

So, the problem can be formulated as follows:

Problem 1. Let $h(\lambda) : (0, \infty) \to \mathbb{R}$ be a given (arbitrary) parametric function and suppose that $X_1, \ldots, X_n$ are iid with density (2.1). Under what conditions on h is it possible to find an estimator $T = T(X_1, \ldots, X_n)$ such that

(i) $\mathbb{E}_\lambda T = h(\lambda)$ for all $\lambda > 0$, and

(ii) $\mathbb{E}_\lambda T^2 < \infty$ for all $\lambda > 0$?

And, in case that such a T exists, how can we obtain the best possible estimator for h?

An estimator satisfying condition 1 is called *unbiased*; as we shall see, unbiasedness restricts the class of possible estimators in such a way that the family of permitted parametric functions h is quite narrow. Condition 2 means that $T \in \cap_{\lambda>0} L^2(\mu_n(\lambda))$, where $\mu_n(\lambda)$ is the product probability measure of $(X_1, \ldots, X_n)$ on $[0, \infty)^n$. Then, provided $\mathbb{E}_\lambda T = h(\lambda)$, the quantity $\mathbb{E}_\lambda (T - h(\lambda))^2$ can be written as $\mathrm{Var}_\lambda T = \mathbb{E}_\lambda T^2 - h(\lambda)^2$, and it is called the *variance* of the estimator T. Even if T is not unbiased, the quantity $\mathbb{E}_\lambda (T - h(\lambda))^2$ is called MSE (mean squared error), and it is the most important measure of closeness between a point estimator $T(X_1, \ldots, X_n)$ and a parametric function $h(\lambda)$, traditionally used in statistics for a long time. The subscript λ in $\mathbb{E}_\lambda$ and Var_λ denotes that the true probability measure of the X_i's is as in (2.1).

It is clear that, if we restrict ourselves to the class of unbiased estimators, those with smaller variance are preferable. In the plausible case that we can pick an estimator T^* satisfying

(i) $\mathbb{E}_\lambda T^* = h(\lambda)$ for all $\lambda > 0$,

(ii) $\mathrm{Var}_\lambda T^* < \infty$ for all $\lambda > 0$, and

(iii) for any unbiased estimator T and for all $\lambda > 0$, $\mathrm{Var}_\lambda T^* \le \mathrm{Var}_\lambda T$, it follows that this is the best we can do. Such an estimator T^* is then called *uniformly minimum variance unbiased estimator* (UMVUE for short), and this is what we could name as *best*. In order to be able to obtain the UMVUE it is necessary and sufficient that the class

$$\mathcal{T}_h = \{T : T \text{ is an unbiased estimator for } h(\lambda) \text{ with finite variance (for all } \lambda > 0)\}$$

is nonempty. This follows from one of the most fundamental result in parametric inference, adapted to the present particular case of Gamma distributions with a known. Indeed, the following is true; see Lehmann and Gasella (1998).

Theorem 2.1. (Rao-Blackwell / Lehmann-Scheffé). *Let $X_1, \ldots, X_n$ be a random sample from* (2.1) *with $\lambda > 0$ unknown and $a > 0$ known. Let also $h : (0, \infty) \to \mathbb{R}$ be a parametric function, and suppose that $\mathcal{T}_h$ is nonempty. Set $X = X_1 + \cdots + X_n$. Then,*

(i) *The conditional probability distribution of $(X_1, \ldots, X_n)$ given X is independent of λ.*

(ii) *For any $T \in \mathcal{T}_h$, the* (unique w.p. 1) *UMVUE is given by the conditional expectation*

$$T^*(X) := \mathbb{E}\Big(T(X_1, \ldots, X_n) \;\Big|\; X\Big).$$

(iii) *Equivalently, the UMVUE of $h(\lambda)$ is the unique* (w.p. 1) *unbiased estimator in $\mathcal{T}_h$ which is a function of X, $u = u(X)$. Hence, $u(X) = \mathbb{E}\Big(T(X_1, \ldots, X_n) \;\Big|\; X\Big) = T^*(X)$.*

Remark 2.1. It is well-known that the distribution of $X = X_1 + \cdots + X_n$ is $\Gamma(na, \lambda)$. In view of Theorem 2.1, and substituting a for $na > 0$, Problem 1 reduces to an equivalent, much simper, reformulation, as follows.

Problem 2. Let $h(\lambda) : (0, \infty) \to \mathbb{R}$ be a given (arbitrary) parametric function and suppose that X is a random variable having probability density (2.1), with $a > 0$ fixed and known, and $\lambda > 0$ an unknown parameter. Under what conditions on h does the UMVUE $u = u(X)$ of $h(\lambda)$ exists for all λ? And, in case that it exists, how can we obtain its form?

Since, by definition, $\mathbb{E}_\lambda \psi(X) = \int_0^\infty f_\lambda(x)\psi(x)\mathrm{d}x$ for arbitrary measurable ψ, the imposed condition of a finite second moment on u *for all* λ implies that

$$\int_0^\infty x^{a-1} \exp(-\lambda x) u(x)^2 \mathrm{d}x < \infty. \tag{2.2}$$

In other words, $u \in L^2(\lambda)$ for all $\lambda > 0$, where $L^2(\lambda)$ is the Lebesgue space of functions $u : (0, \infty) \to \mathbb{R}$ satisfying (2.2). Thus, it is reasonable to define

$$L_0^2 := \textstyle\bigcap_{\lambda > 0} L^2(\lambda). \tag{2.3}$$

We can rewrite the unbiasedness restriction $\mathbb{E}_\lambda u(X) = h(\lambda)$ as

$$\frac{\Gamma(a)h(\lambda)}{\lambda^a} = \int_0^\infty x^{a-1} \exp(-\lambda x) u(x) \mathrm{d}x, \quad \lambda > 0. \tag{2.4}$$

It is then obvious that the rhs of (2.4) defines a holomorphic function in the right half-plane $\mathcal{C}^+ = \{\lambda \in \mathcal{C} : \mathrm{Re}(\lambda) > 0\}$ whenever $u \in L_0^2$. This means that the function $\lambda \to$

$\Gamma(a)\lambda^{-a}h(\lambda)$ is holomorphic, and hence, $h(\lambda)$ must be holomorphic in $\mathcal{C}^+$. This already imposes a serious restriction to the allowable parametric functions, e.g., it is necessary that $h \in C^\infty(0,\infty)$; in fact, the analytic extension of h should have no singularities in the right half-plane. As a simple example, for the $C^\infty(0,\infty)$ parametric function $h(\lambda) = 1/(\lambda^2-2\lambda+2)$, no unbiased estimator exist (for all $\lambda > 0$), because of the poles $1 \pm i$ of h. However, regarding Problem 2, the analyticity of h is not sufficient to provide a positive answer. To see this, it suffices to observe that for $u \in L_0^2$, $\int_0^\infty x^{a-1}\exp(-\lambda x)u(x)\mathrm{d}x \to 0$ as $\lambda \to +\infty$, by dominated convergence. Then, any holomorphic function h that growths faster than e^λ at infinity, e.g. $h(\lambda) = \exp(\lambda^2)$, cannot be written as the expectation of some $u \in L_0^2$; see (2.4).

2.2 Results

We are now in a position to state the main results.

Theorem 2.2. *Assume that X is a random variable with probability density f_λ as in* (2.1)*, with $\lambda > 0$ unknown. For a given parametric function $h(\lambda)$, its UMVUE $u(X)$ exists in L_0^2 if and only if the following two conditions are satisfied.*
(1) *The function h can be extended to an holomorphic function in $\mathcal{C}^+$, and*
(2) *For any $\lambda > 0$,*

$$\sum_{n=0}^{\infty} \beta_n(\lambda)^2 < \infty, \tag{2.5}$$

where

$$\beta_n(\lambda) = \frac{(-1)^n}{\sqrt{n![a]_n}}\left(\lambda \frac{\mathrm{d}^n}{\mathrm{d}\lambda^n}\left[\lambda^{n-1}h(\lambda)\right]\right); \tag{2.6}$$

here, $[a]_n = \prod_{j=0}^{n-1}(a+j) = \Gamma(a+n)/\Gamma(a)$ denotes the ascending factorial (Pochhammer symbol).

Theorem 2.3. *Let h be a parametric function satisfying* (1) *and* (2) *of Theorem* 2.2. *For fixed $\lambda > 0$ define the function*

$$H_\lambda(y) := h\Big(\frac{\lambda}{1-y}\Big), \quad |y| < 1. \tag{2.7}$$

Then, an alternative simplified form of the constants $\beta_n(\lambda)$ in (2.6) *is given by*

$$\beta_n(\lambda) = \frac{(-1)^n H_\lambda^{(n)}(0)}{\sqrt{n![a]_n}}. \tag{2.8}$$

Theorem 2.4. *Assume that* (1) *and* (2) *of Theorem* 2.2 *are satisfied and fix $\lambda_0 > 0$. Then, the function $u(x)$ for which $u(X)$ is the UMVUE of $h(\lambda)$ is given by*

$$u(x) = \sum_{n=0}^{\infty} \beta_n(\lambda_0) q_{n;\lambda_0}(x) \tag{2.9}$$

where $\{q_{n;\lambda_0}(x)\}_{n=0}^\infty$ is the complete orthonormal polynomial system corresponding to the weight function f_{λ_0}, with the convention that each $q_{n;\lambda_0}$ is of degree n and with strictly *positive leading coefficient. The series converges a.e. on $(0,\infty)$ and in $L^2(\lambda)$ for every $\lambda \ge \lambda_0$, and the resulting function $u(x)$, given by* (2.9)*, is independent of the choice of λ_0. Furthermore, for any $\lambda > 0$, the variance of the UMVUE is given by*

$$\mathrm{Var}_\lambda u(X) = \sum_{n=1}^{\infty} \beta_n(\lambda)^2, \tag{2.10}$$

where the constants $\beta_n(\lambda)$ are completely determined from $h(\lambda)$; see (2.6) *or* (2.8).

Example 1. We compare the expression (2.10) with the classical information inequality, namely, the famous Cramér-Rao (CR) lower bound (LB$_{CR}$). Since, as is well-known, the regularity conditions are satisfied for f_λ, the bound states that for a random sample $X_1,\ldots,X_n$ (of size n) from f_λ, and for any unbiased estimator $T = T(X_1,\ldots,X_n)$ of $h(\lambda)$, the inequality $\mathrm{Var}_\lambda T \ge h'(\lambda)^2/(nI(\lambda)) :=$ LB$_{CR}$ is satisfied; here, $I(\lambda)$ is the Fisher information, defined as

$$I(\lambda) := \mathbb{E}_\lambda\left[\left(\frac{\partial}{\partial\lambda}\log f_\lambda(X_1)\right)^2\right] = \frac{a}{\lambda^2}.$$

Thus, the CR-bound reads as $\mathrm{Var}_\lambda T \ge \lambda^2 h'(\lambda)^2/(na)$. On the other hand, the series expansion (2.10) (applied with na in place of a; see Problems 1 and 2) yields

$$\mathrm{Var}_\lambda u(X) = \sum_{m=1}^{\infty} \frac{\lambda^2}{m![na]_m}\left(\frac{\mathrm{d}^m}{\mathrm{d}\lambda^m}\Big[\lambda^{m-1}h(\lambda)\Big]\right)^2.$$

Since $u(X)$ is the UMVUE and thus, $\mathrm{Var}_\lambda T \ge \mathrm{Var}_\lambda u(X)$ for any unbiased estimator T, it is clear that the CR-bound is implied by the preceding series, on just keeping its first term.

2.3 Proofs

We first state some auxiliary lemmas.

Lemma 2.2. *For $x > 0$, $a > 0$ and $\lambda > 0$,*

$$\frac{\mathrm{d}^n}{\mathrm{d}x^n}\Big[x^n f_\lambda(x)\Big] = \lambda\frac{\mathrm{d}^n}{\mathrm{d}\lambda^n}\Big[\lambda^{n-1} f_\lambda(x)\Big], \quad n = 0,1,2,\ldots\,. \tag{2.11}$$

Proof. By Leibnitz formula and (2.1) it is easily seen that both sides of (2.11) are equal to

$$\Gamma(a+n) f_\lambda(x)\sum_{k=0}^{n}(-1)^k\binom{n}{k}\frac{(\lambda x)^k}{\Gamma(a+k)}.$$

Lemma 2.3. (Rodrigues' formula; see Afendras and Papadatos (2015)). *For $x > 0$, $a > 0$ and $\lambda > 0$,*

$$\frac{\mathrm{d}^n}{\mathrm{d}x^n}\Big[x^n f_\lambda(x)\Big] = (-1)^n\sqrt{[a]_n n!}\, f_\lambda(x) q_{n;\lambda}(x), \quad n = 0,1,2,\ldots, \tag{2.12}$$

where $\{q_{n;\lambda}(x)\}_{n=0}^{\infty}$ is the complete orthonormal system with respect to f_λ, standardized so that $q_{n;\lambda}$ has degree n and positive leading coefficient. The polynomials $q_{n;\lambda}$ satisfy the orthogonality condition

$$\mathbb{E}_\lambda\Big[q_{n;\lambda}(X)q_{m;\lambda}(X)\Big] = \int_0^\infty f_\lambda(x) q_{n;\lambda}(x) q_{m;\lambda}(x)\,\mathrm{d}x = \begin{cases} 1 & \text{if } n = m; \\ 0 & \text{if } n \ne m. \end{cases}$$

One important observation is that, as (2.12) and (2.11) show, we may produce the orthonormal set $q_{n;\lambda}$ by differentiate w.r. to the parameter λ, instead of x.; more precisely,

$$q_{n;\lambda}(x) = \frac{(-1)^n}{\sqrt{n![a]_n} f_\lambda(x)} \frac{\mathrm{d}^n}{\mathrm{d}x^n}\Big[x^n f_\lambda(x)\Big] = \frac{(-1)^n}{\sqrt{n![a]_n} f_\lambda(x)} \left(\lambda \frac{\mathrm{d}^n}{\mathrm{d}\lambda^n}\Big[\lambda^{n-1} f_\lambda(x)\Big]\right). \tag{2.13}$$

Thus, (2.13) obtains the following

Corollary 2.4. *For $x>0$, $a>0$, $\lambda>0$ and $n \in \{0,1,\ldots\}$,*

$$q_{n;\lambda}(x) f_\lambda(x) = \frac{(-1)^n}{\sqrt{n![a]_n}} \left(\lambda \frac{d^n}{d\lambda^n}\Big[\lambda^{n-1} f_\lambda(x)\Big]\right). \tag{2.14}$$

We now proceed to verify the results of Theorems 2.2–2.4.

Assume first that the UMVUE of $h(\lambda)$ is $u(X)$, and suppose that it has finite variance for all $\lambda > 0$. Multiplying the equation $\mathbb{E}_\lambda u(X) = h(\lambda)$ by λ^{n-1} and then taking n derivatives w.t. to λ, we subsequently obtain

$$\begin{aligned}
h(\lambda) &= \int_0^\infty f_\lambda(x)u(x)\mathrm{d}x,\\
\lambda^{n-1}h(\lambda) &= \int_0^\infty \lambda^{n-1} f_\lambda(x)u(x)\mathrm{d}x,\\
\frac{\mathrm{d}^n}{\mathrm{d}\lambda^n}\Big[\lambda^{n-1}h(\lambda)\Big] &= \int_0^\infty \left(\frac{\mathrm{d}^n}{\mathrm{d}\lambda^n}\Big[\lambda^{n-1} f_\lambda(x)\Big]\right) u(x)\mathrm{d}x,\\
\lambda\frac{\mathrm{d}^n}{\mathrm{d}\lambda^n}\Big[\lambda^{n-1}h(\lambda)\Big] &= \int_0^\infty \left(\lambda\frac{\mathrm{d}^n}{\mathrm{d}\lambda^n}\Big[\lambda^{n-1} f_\lambda(x)\Big]\right) u(x)\mathrm{d}x,\\
\frac{(-1)^n}{\sqrt{n![a]_n}}\left(\lambda\frac{\mathrm{d}^n}{\mathrm{d}\lambda^n}\Big[\lambda^{n-1}h(\lambda)\Big]\right) &= \int_0^\infty q_{n;\lambda}(x) f_\lambda(x)u(x)\mathrm{d}x;
\end{aligned} \tag{2.15}$$

note that the differentiation can be passed under the integral sign, due to the assumed (squared) integrability of u with respect to f_λ for all $\lambda > 0$. We conclude from (2.15) that the constants $\beta_n(\lambda)$ of (2.6) are the Fourier coefficients of u with respect to the orthonormal polynomial system $\{q_{n;\lambda}\}_{n=0}^\infty$, corresponding to the weight function f_λ. It should be noticed that the orthonormal polynomial system corresponding to a probability measure (having finite moments of any order) is unique, apart from a possible multiplication of each polynomial by ±1. Moreover, since our system $\{q_{n;\lambda}\}_{n=0}^\infty$ is complete in $L^2(\lambda)$, see Afendras *et al* (2011), Parseval's identity yields

$$\mathbb{E}_\lambda u(X)^2 = \int_0^\infty f_\lambda(x)u(x)^2\mathrm{d}x = \sum_{n=0}^\infty \beta_n(\lambda)^2 < \infty.$$

Thus, assuming that $u \in L_0^2$, and since $\beta_0(\lambda) = \mathbb{E}_\lambda u(X) = h(\lambda)$, we get

$$\mathrm{Var}_\lambda u(X) = \sum_{n=1}^\infty \beta_n(\lambda)^2, \quad \text{for all } \lambda > 0.$$

Conversely, assume that h is holomorphic in $\mathcal{C}^+$ and that the series in (2.6) is finite for all $\lambda > 0$. Then we may define the function $u(x;\lambda)$ as

$$u(x;\lambda) := \sum_{n=0}^\infty \beta_n(\lambda) q_{n;\lambda}(x), \quad x > 0, \tag{2.16}$$

where, by Riesz-Fisher, the series converges in $L^2(\lambda)$, that is,

$$\int_0^\infty \Big(u_N(x;\lambda) - u(x;\lambda)\Big)^2 f_\lambda(x)\mathrm{d}x \to 0, \quad N \to \infty,$$

with $u_N(x;\lambda) = \sum_{n=0}^N \beta_n(\lambda) q_{n;\lambda}(x)$. It remains to show that the limiting function $u(x;\lambda)$ does not depend on λ, and that it is the unique UMVUE of $h(\lambda)$. To this end, choose a fixed $\lambda_0 > 0$ with $\lambda_0 < \lambda$ and write

$$u_N(x) = \sum_{n=0}^N \beta_n(\lambda_0) q_{n;\lambda_0}(x), \quad x > 0, \; N = 0, 1, 2, \dots \; . \tag{2.17}$$

From the convergence of the series (2.5) (with $\lambda = \lambda_0$) it is easily seen that $u_N(x)$ is Cauchy $L^2(\lambda_0)$, and hence, it converges (in the norm of $L^2(\lambda_0)$) to a function $u(x) \in L^2(\lambda_0)$. Moreover, is easy to check that for any $\lambda \geq \lambda_0$, we can find a constant $C_\lambda = C(\lambda, \lambda_0)$ such that $f_\lambda(x) \leq C_\lambda f_{\lambda_0}(x)$ for all $x > 0$. This implies that u_N is also Cauchy $L^2(\lambda)$ for any fixed $\lambda \geq \lambda_0$; indeed, if $\epsilon > 0$ is arbitrary, we can find $N(\epsilon)$ such that $\int_0^\infty (u_{N_1}(x) - u_{N_2}(x))^2 f_{\lambda_0}(x)\mathrm{d}x < \epsilon / C_\lambda$ for all $N_1, N_2 > N(\epsilon)$ and, then,

$$\int_0^\infty (u_{N_1}(x) - u_{N_2}(x))^2 f_\lambda(x)\mathrm{d}x \leq C_\lambda \int_0^\infty (u_{N_1}(x) - u_{N_2}(x))^2 f_{\lambda_0}(x)\mathrm{d}x < \epsilon.$$

The preceding argument verifies that the limiting function u, defined as the $L^2(\lambda_0)$-limit of the sequence in (2.17), belongs to $L^2(\lambda)$ for all $\lambda \geq \lambda_0$, in symbols, $u(x) \in \bigcap_{\lambda \geq \lambda_0} L^2(\lambda)$. From the orthogonality of the polynomials $q_{n;\lambda_0}$ ($n \geq 1$) and $q_{0;\lambda_0} \equiv 1$ we immediately see that $\mathbb{E}_{\lambda_0} u_N(X) = \beta_0(\lambda_0) = h(\lambda_0)$, and clearly, this is also true for u, i.e., $\mathbb{E}_{\lambda_0} u(X) = h(\lambda_0)$. However, the situation is different when $\lambda \neq \lambda_0$, that is, the expectation of $u_N(X)$ w.r. to f_λ varies with both N and λ. More precisely, since $q_{0;\lambda}(x) \equiv 1$,

$$\mathbb{E}_\lambda u_N(X) = h(\lambda_0) + \sum_{n=1}^N \beta_n(\lambda_0) \mathbb{E}_\lambda q_{n;\lambda_0}(X), \quad N = 1, 2, \dots \quad \lambda > 0.$$

On the other hand, we have shown that for $\lambda \geq \lambda_0$, $\mathbb{E}_\lambda \Big(u_N(X) - u(X)\Big)^2 \to 0$, so that, by the Cauchy-Schwarz inequality,

$$\Big|\mathbb{E}_\lambda u_N(X) - \mathbb{E}_\lambda u(X)\Big| \leq \mathbb{E}_\lambda \big|u_N(X) - u(X)\big| \leq \Big(\mathbb{E}_\lambda \big|u_N(X) - u(X)\big|^2\Big)^{1/2} \to 0.$$

It follows that $\mathbb{E}_\lambda u(X) = \lim_N \mathbb{E}_\lambda u_N(X)$. Hence, the expectation of $u(X)$ w.r. to f_λ can be obtained as the limit of the expectations of $u_N(X)$ (w.r. to f_λ). Next, we see that the calculation of $\mathbb{E}_\lambda u_N(X)$ requires evaluation of the expectations $\mathbb{E}_\lambda q_{n;\lambda_0}(X)$, that is, integrals of the polynomials $q_{n;\lambda_0}(x)$ w.r. to a different weight function (f_λ instead of f_{λ_0}), under which these polynomials are no longer orthogonal.

In order to calculate $\mathbb{E}_\lambda q_{n;\lambda_0}(X)$ we proceed as follows. We have

$$\begin{aligned} \mathbb{E}_\lambda q_{n;\lambda_0}(X) &= \int_0^\infty \frac{f_\lambda(x)}{f_{\lambda_0}(x)} f_{\lambda_0}(x) q_{n;\lambda_0}(x)\mathrm{d}x \\ &= \Big(\frac{\lambda}{\lambda_0}\Big)^a \int_0^\infty f_{\lambda_0}(x) \exp\Big(-(\lambda - \lambda_0)x\Big) q_{n;\lambda_0}(x)\mathrm{d}x. \end{aligned}$$

The last integral can be viewed as the n-th Fourier coefficient of the bounded $C^\infty(0,\infty)$ function $w(x) := \exp\Big(-(\lambda-\lambda_0)x\Big)$, $x>0$, with respect to the corresponding orthonormal polynomial system $\{q_{n;\lambda_0}\}_{n=0}^\infty$ in $L^2(\lambda_0)$. On the other hand, it is known that the same Fourier coefficients can be conveniently obtained by using the identity (see Afendras and Papadatos (2015), Afendras *et al* (2011))

$$\mathbb{E}_{\lambda_0}\Big[q_{n;\lambda_0}(X)w(X)\Big] = \frac{1}{\sqrt{n![a]_n}}\mathbb{E}_{\lambda_0}\Big[X^n w^{(n)}(X)\Big],$$

provided $\mathbb{E}_{\lambda_0}\Big[X^n\Big(w^{(n)}(X)\Big)^2\Big]<\infty$. Since $w^{(n)}(x)=(\lambda_0-\lambda)^n\exp\Big(-(\lambda-\lambda_0)x\Big)$ is a bounded function of x, because $\lambda\ge\lambda_0$, we can apply the preceding formulae to deduce

$$\mathbb{E}_\lambda q_{n;\lambda_0}(X) = \Big(\frac{\lambda}{\lambda_0}\Big)^a\frac{(\lambda_0-\lambda)^n}{\sqrt{n![a]_n}}\mathbb{E}_{\lambda_0}\Big[X^n\exp\Big(-(\lambda-\lambda_0)X\Big)\Big].$$

A straightforward computation now yields

$$\mathbb{E}_{\lambda_0}\Big[X^n\exp\Big(-(\lambda-\lambda_0)X\Big)\Big] = \frac{\lambda_0^a}{\Gamma(a)}\int_0^\infty x^{n+a-1}\exp(-\lambda x)\mathrm{d}x = [a]_n\frac{\lambda_0^a}{\lambda^{n+a}},$$

and thus,

$$\mathbb{E}_\lambda q_{n;\lambda_0}(X) = (-1)^n\sqrt{\frac{[a]_n}{n!}}\Big(1-\frac{\lambda_0}{\lambda}\Big)^n.$$

Recalling that $\beta_n(\lambda_0)$ is given by (2.6) with $\lambda=\lambda_0$, we have

$$\begin{aligned}\mathbb{E}_\lambda u_N(X) &= \sum_{n=0}^N\frac{(-1)^n\lambda_0}{\sqrt{n![a]_n}}\frac{\mathrm{d}^n}{\mathrm{d}\lambda^n}\Big[\lambda^{n-1}h(\lambda)\Big]\Big|_{\lambda=\lambda_0}\left((-1)^n\sqrt{\frac{[a]_n}{n!}}\Big(1-\frac{\lambda_0}{\lambda}\Big)^n\right)\\ &= \sum_{n=0}^N\frac{1}{n!}\Big(1-\frac{\lambda_0}{\lambda}\Big)^n\left\{\lambda_0\frac{\mathrm{d}^n}{\mathrm{d}\lambda^n}\Big[\lambda^{n-1}h(\lambda)\Big]\Big|_{\lambda=\lambda_0}\right\}.\end{aligned}\tag{2.18}$$

Though the preceding formula appears to be quite complicated at a first glance (e.g., it seems that it is not an easy task to obtain its limiting value as $N\to\infty$), this is not the case. In fact, (2.18) represents a Taylor development around $y=0$ of the function $H_{\lambda_0}(y) := h\Big(\frac{\lambda_0}{1-y}\Big)$, $|y|<1$. Recall that $h(\lambda)$ has been assumed to be holomorphic in $\mathrm{Re}(\lambda)>0$, so that $H_{\lambda_0}(y)$ is analytic in the open disc $|y|<1$. Writing $H_{\lambda_0}^{(n)}(y)$ for $\frac{\mathrm{d}^n}{\mathrm{d}y^n}H_{\lambda_0}(y)$, we shall verify below the equality

$$H_{\lambda_0}^{(n)}(0) = \left\{\lambda_0\frac{\mathrm{d}^n}{\mathrm{d}\lambda^n}\Big[\lambda^{n-1}h(\lambda)\Big]\Big|_{\lambda=\lambda_0}\right\},\quad n=0,1,\ldots\,. \tag{2.19}$$

Assuming for a while that (2.19) is valid, and substituting it to (2.18), we obtain the simple formula

$$\mathbb{E}_\lambda u_N(X) = \sum_{n=0}^N\frac{H_{\lambda_0}^{(n)}(0)}{n!}\Big(1-\frac{\lambda_0}{\lambda}\Big)^n.$$

Since $|1-\lambda_0/\lambda|<1$ (for $\lambda>\lambda_0/2$), we conclude from Taylor's theorem that $\mathbb{E}_\lambda u_N(X)\to H_{\lambda_0}(1-\lambda_0/\lambda)=h(\lambda)$. Thus, $\mathbb{E}_\lambda u(X)=\lim_N\mathbb{E}_\lambda u_N(X)=h(\lambda)$, and this verifies that $u(X)$ is the (unique) UMVUE of $h(\lambda)$, for every $\lambda\ge\lambda_0$. [To see uniqueness, repeat the previous construction with λ_1 in place of λ_0. Then, as we showed, the produced estimator $\widetilde{u}(X)$ will satisfy $\mathbb{E}_\lambda \widetilde{u}(X) = h(\lambda) = \mathbb{E}_\lambda u(X)$ for all $\lambda \geq \max\{\lambda_0, \lambda_1\}$, so it must be identical to $u(X)$.] Furthermore, (2.6) shows that u has the same Fourier coefficients as the function $u(x;\lambda)$ defined by (2.16); thus $u(x) = u(x;\lambda)$ is independent of λ, and Parseval's identity yields (2.10). The orthogonal polynomials for the weight function f_λ are called generalized Laguerre (Laguerre when $a = 1$). The a.e. convergence of the Laguerre series expansion of a function $u \in L^2(\lambda)$ is a well-known (Carleson-Hunt type) result that can be found in Mackenhoupt (1970); see also Uspensky (1927) and Stempak (2000).

It remains to show (2.19). Using Leibnitz formula we first calculate

$$\lambda \frac{\mathrm{d}^n}{\mathrm{d}\lambda^n}\Big[\lambda^{n-1} h(\lambda)\Big] = (n-1)! \sum_{k=1}^{n} \binom{n}{k} \frac{\lambda^k h^{(k)}(\lambda)}{(k-1)!}, \quad n = 1, 2, \ldots, \tag{2.20}$$

while the lhs equals to $h(\lambda)$ for $n = 0$. Next, we define $H_\lambda(y) = h\Big(\lambda/(1-y)\Big)$, $|y| < 1$, so that $H_\lambda^{(0)}(y) = H_\lambda(y)$ and $H_\lambda^{(0)}(0) = h(\lambda)$. For $n = 1$, $H_\lambda'(y) = \lambda h'\Big(\lambda/(1-y)\Big)\Big/(1-y)^2$, and $H_\lambda'(0) = \lambda h'(\lambda)$ equals to the sum in the rhs of (2.20) (with $n = 1$). We shall prove, using induction on n, the formula (valid for $\lambda > 0$, $|y| < 1$)

$$H_\lambda^{(n)}(y) = (n-1)! \sum_{k=1}^{n} \binom{n}{k} \frac{\lambda^k h^{(k)}(\lambda/(1-y))}{(k-1)!(1-y)^{n+k}}, \quad n = 1, 2, \ldots, \tag{2.21}$$

which, setting $y = 0$, yields the rhs of (2.20); then, the substitution $\lambda \to \lambda_0$ verifies (2.19). Noting that (2.21) is true for $n = 1$, we assume that it is true for some n. Then,

$$\begin{aligned}
H_\lambda^{(n+1)}(y) &= \frac{\mathrm{d}}{\mathrm{d}y}\left\{(n-1)! \sum_{k=1}^{n} \binom{n}{k} \frac{\lambda^k h^{(k)}(\lambda/(1-y))}{(k-1)!(1-y)^{n+k}}\right\} \\
&= (n-1)! \sum_{k=1}^{n} \binom{n}{k} \frac{\lambda^k}{(k-1)!} \frac{\mathrm{d}}{\mathrm{d}y}\left\{\frac{h^{(k)}(\lambda/(1-y))}{(1-y)^{n+k}}\right\} \\
&= (n-1)! \sum_{k=1}^{n} \binom{n}{k} \frac{\lambda^k}{(k-1)!} \frac{h^{(k+1)}(\lambda/(1-y))}{(1-y)^{n+k}} \frac{\lambda}{(1-y)^2} \\
&\quad + (n-1)! \sum_{k=1}^{n} \binom{n}{k} \frac{\lambda^k}{(k-1)!} \frac{h^{(k)}(\lambda/(1-y))}{(1-y)^{n+k+1}} (n+k) \\
&= (n-1)! \sum_{k=2}^{n+1} (k-1) \binom{n}{k-1} \frac{\lambda^k h^{(k)}(\lambda/(1-y))}{(k-1)!(1-y)^{n+1+k}} \\
&\quad + (n-1)! \sum_{k=1}^{n} (n+k) \binom{n}{k} \frac{\lambda^k h^{(k)}(\lambda/(1-y))}{(k-1)!(1-y)^{n+1+k}} \\
&= (n-1)! \sum_{k=1}^{n+1} \left\{(k-1)\binom{n}{k-1} + (n+k)\binom{n}{k}\right\} \frac{\lambda^k h^{(k)}(\lambda/(1-y))}{(k-1)!(1-y)^{n+1+k}},
\end{aligned}$$

where the last equality follows from $\binom{n}{k} = 0$ for $k = n+1$ and $(k-1)\binom{n}{k-1} = 0$ for

$k = 1$. It is now obvious that

$$
\begin{aligned}
(k-1)\binom{n}{k-1} + (n+k)\binom{n}{k} &= \frac{(k-1)n!}{(k-1)!(n-k+1)!} + \frac{(n+k)n!}{k!(n-k)!} \\
&= \Big\{k(k-1) + (n+k)(n+1-k)\Big\}\frac{n!}{k!(n+1-k)!} \\
&= n(n+1)\frac{n!}{k!(n+1-k)!} \\
&= n\binom{n+1}{k}.
\end{aligned}
$$

This shows that (2.21) holds with $n+1$ in place of n, and concludes the inductional argument.

3 A novel inversion formula of the Laplace transform

The results of Section 2 apply to the particular case where $a = 1$, i.e., when X follows the exponential distribution with parameter $\lambda > 0$, $\text{Exp}(\lambda)$, with probability density

$$f_\lambda(x) = \lambda \exp(-\lambda x), \quad x > 0. \tag{3.1}$$

In this case, Lemma 2.3 produces the corresponding orthonormal polynomial system, namely,

$$q_{n;\lambda}(x) = \sum_{k=0}^{n} (-1)^{n-k}\binom{n}{k}\frac{(\lambda x)^k}{k!}.$$

The preceding polynomials are functions of λx (this is also true for $a \neq 1$, since it is easily seen that $q_{n;\lambda}(x) = q_{n;1}(\lambda x)$). Hence, it is convenient to define $p_n(x) = q_{n;1}(x)$, so that $q_{n;\lambda}(x) = p_n(\lambda x)$. Then, the polynomial system $\{p_n(x)\}_{n=0}^{\infty}$ is the complete orthonormal system corresponding to f_1 (i.e., $\text{Exp}(1)$), that is,

$$\mathbb{E}\Big[p_n(X)p_m(X)\Big] = \int_0^\infty e^{-x} p_n(x)p_m(x)\mathrm{d}x = \begin{cases} 1 & \text{if} \quad n = m, \\ 0 & \text{if} \quad n \neq m, \end{cases}$$

where $\mathbb{E}$ stands for $\mathbb{E}_1$. Traditionally, the polynomials $L_n(x) = (-1)^n p_n(x)$ (with alternating leading coefficients) are called *Laguerre polynomials*, see (1.6), and they are also orthonormal w.r. to $f_1(x) = e^{-x}$, $x > 0$.

Consider now Problem 2 with $a = 1$. This reduces in finding the function

$$u \in L_0^2 := \textstyle\bigcap_{\lambda>0} L^2\Big((0,\infty), e^{-\lambda x}\Big)$$

for which

$$\mathbb{E}_\lambda u(X) := \int_0^\infty \lambda \exp(-\lambda x)u(x)\mathrm{d}x = h(\lambda), \quad \lambda > 0,$$

provided that $h(\lambda)$ allows such a construction. Theorem 2.2 provides a necessary and sufficient condition on h, namely, $h(\lambda)$ is holomorphic for $\lambda \in \mathcal{C}^+ = \{\lambda \in \mathcal{C} : \text{Re}(\lambda) > 0\}$ satisfying

$$\sum_{n=0}^{\infty}\left(\frac{(-1)^n}{n!}\left(\lambda\frac{\mathrm{d}^n}{\mathrm{d}\lambda^n}\Big[\lambda^{n-1}h(\lambda)\Big]\right)\right)^2 < \infty, \quad \lambda > 0. \tag{3.2}$$

In view of Theorem 2.3, the preceding condition can be rewritten as

$$\sum_{n=0}^{\infty}\left(\frac{(-1)^n H_\lambda^{(n)}(0)}{n!}\right)^2 < \infty, \quad \lambda > 0, \tag{3.3}$$

where $H_\lambda(y) = h\Big(\lambda/(1-y)\Big)$, $|y| < 1$.

It is obvious that the equation $\mathbb{E}_\lambda u(X) = h(\lambda)$ can be written in terms of the Laplace transform of u, (1.1), as

$$\lambda\phi(\lambda) = \int_0^\infty \lambda \exp(-\lambda x) u(x) \mathrm{d}x = \mathbb{E}_\lambda u(X) = h(\lambda).$$

Hence, given the (holomorphic in $\mathcal{C}^+$) Laplace transform ϕ, one can check the validity of either (3.2) or (3.3) for $h(\lambda) := \lambda\phi(\lambda)$, in order to ensure that the inverse function $u(x)$ of $\phi(\lambda)$ exists in L_0^2; if this is the case, then Theorem (2.4) applies and u is obtained as a Laguerre polynomial series with constants derived from the derivatives of ϕ.

Translating Theorems 2.2–2.4 to the Laplace-transform case, we obtain the following

Theorem 3.1. (A) *Assume that $\phi(\lambda)$ is an holomorphic function of $\lambda \in \mathcal{C}^+$, such that*

$$\sum_{n=0}^{\infty}\left(\frac{1}{n!}\left(\lambda\frac{\mathrm{d}^n}{\mathrm{d}\lambda^n}\Big[\lambda^n\phi(\lambda)\Big]\right)\right)^2 < \infty, \quad \lambda > 0, \tag{3.4}$$

or, equivalently,

$$\sum_{n=0}^{\infty}\left(\frac{\Phi_\lambda^{(n)}(0)}{n!}\right)^2 < \infty, \quad \lambda > 0, \tag{3.5}$$

where

$$\Phi_\lambda(y) = \frac{\lambda}{1-y}\,\phi\Big(\frac{\lambda}{1-y}\Big), \quad |y| < 1. \tag{3.6}$$

Then, ϕ is the Laplace transform of a function $u \in L_0^2$. Moreover, for every fixed $\lambda_0 > 0$, the inverse Laplace transform, u, is given by

$$u(x) = \sum_{n=0}^{\infty}\frac{\Phi_{\lambda_0}^{(n)}(0)}{n!}\,L_n(\lambda_0 x), \tag{3.7}$$

where the Laguerre polynomials L_n are given by (1.6). *The series converges a.e. and in $L^2\big(\mathbb{R}_+, e^{-\lambda x}\big)$ for every $\lambda \geq \lambda_0$, and the sum of the series does not dependent on the particular choice of λ_0.*
(B) *If ϕ is the Laplace transform of a function $u \in L_0^2$ then ϕ is holomorphic in $\mathcal{C}^+$ and satisfies* (3.4) (equivalently, (3.5)).

Since the choice of λ_0 does not affect the validity of (3.7), we may set $\lambda_0 = 1$. Then, the function Φ_λ in (3.6) reduces to $\Phi_1(y) = (1-y)^{-1}\phi\big((1-y)^{-1}\big) = \Phi(y)$, say, and from (3.7) we obtain the (Taylor-like) Laplace inversion formula

$$u(x) = \sum_{n=0}^{\infty}\frac{\Phi^{(n)}(0)}{n!}\,L_n(x), \quad \text{where } \Phi(y) = \frac{1}{1-y}\,\phi\Big(\frac{1}{1-y}\Big), \tag{3.8}$$

which is valid almost everywhere in $(0, \infty)$.

At this point we note that all inversion formulae of $\phi(\lambda)$ provide approximating functions for $u(x)$ in some sense. For instance, (3.8) says that

$$u_N(x) := \sum_{n=0}^{N} \frac{\Phi^{(n)}(0)}{n!} L_n(x) \to u(x), \quad \text{a.e.,} \tag{3.9}$$

while (1.4) can be written in our case as

$$w_N(x) := \frac{1}{2\pi i} \int_{1-iN}^{1+iN} \exp(sx)\phi(s)\mathrm{d}s \to u(x), \quad \text{a.e.,}$$

and (1.5) reads as

$$v_N(x) := \frac{(-1)^N}{N!} \left(\frac{N}{x}\right)^{N+1} \phi^{(N)}\left(\frac{N}{x}\right) \to u(x) \ \text{ at continuity points } x \text{ of } u(x).$$

Hence, it would be desirable to compare the degree of approximation of the preceding formulae; however, this is beyond the scope of the present work. We merely point out a possible advantage of the new inversion formula: The approximating functions u_N in (3.9) are polynomials, and the formula becomes exact for any polynomial u when $N \geq \deg(u)$.

Acknowledgement. I would like to cordially thank A. Giannopoulos for organizing this meeting in memory of our friend and colleague Dimitris, and also for providing me bibliographic material regarding the a.e. convergence of Laguerre series.

References

[1] Afendras, G.; Papadatos, N. (2014). Strenthened Chernoff-type variance bounds. *Bernoulli*, vol. 20, 245–264.

[2] Afendras, G.; Papadatos, N. (2015). Integrated Pearson family and orthogonality of the Rodrigues polynomials: A review including new results and an alternative classification of the Pearson system. *Applicationes Mathematicae*, vol. 42, 231–267.

[3] Afendras, G.; Papadatos, N.; Papathanasiou, V. (2011). An extended Stein-type covariance identity for the Pearson family, with applications to lower variance bounds. *Bernoulli*, vol. 17, 507–529.

[4] Billingsley, P. (1995). *Probability and Measure*, 3rd ed. Wiley, N.Y.

[5] Boas, M. (1983). *Mathematical Methods in the Physical Sciences*, 2nd ed. Wiley, N.Y.

[6] Cohen, A.M. (2007). Inversion formulae and practical results. In *Numerical Methods for Laplace Transform Inversion*, Numerical Mathods and Algorithms, vol. 5.

[7] Lehmann, E.L.; Gasella, G. (1998). *Theory of Point Estimation*, 2nd. ed. Springer, N.Y.

[8] Lerch, M. (1903). Sur un point de la théorie des fonctions génératrices d'Abel. *Acta Mathematica*, vol. 27, 339–351.

[9] Mackenhoupt, B. (1970). Equiconvergence and almost everywhere convergence of Hermite and Laguerre series. *SIAM J. Math. Anal.*, vol. 3, 295–321.

[10] Post, E. (1930). Generalized differentiation. *Transactions of the Americal Mathematical Society*, vol. 32, 723–781.

[11] Schilling, R.L.; Song R.; Vondraček, Z. (2012). *Bernstein Functions: Theory and Applications*, 2nd ed. De Gruyter Studies Math., vol. 37, De Gruyter, Berlin.

[12] Stempak, K. (2000). Divergent Laguerre series. *Proceeding of the American Mathematical Society*, vol. 129, 1123–1126.

[13] Uspensky, J.V. (1927). On the development of arbitrary functions in series of Laguerre's and Hermite's polynomials. *Annals of Mathematics*, vol. 28, 593–619.

[14] Widder, D.V. (1946). *The Laplace Transform*. Princeton University Press.

Nickos Papadatos: National and Kapodistrian University of Athens, Department of Mathematics, Panepistemiopolis, 157 84, Athens, Greece.

E-mail: npapadat@math.uoa.gr

Systolic inequality and width of metric spaces

Panos Papasoglu

In memory of Dimitris Gatzouras

Abstract

The systole $\mathrm{sys}(M^n)$ of a Riemannian manifold M^n is the length of the shortest non-contractible closed curve on M^n. Lowener showed in 1949 that for the torus T (with any metric) one has $(\mathrm{sys}\, T)^2 \leq \frac{2}{\sqrt{3}}\,\mathrm{area}\, T$. Gromov generalized this result to all aspherical manifolds in 1983. We give here another approach to Gromov's systolic inequality based on a conjecture of Guth about the Uryson width of metric spaces.

1 Some History

The problem of the existence of closed geodesics in closed Riemannian manifolds M^n was raised by Poincaré in 1905. It is relatively easy to see (modulo technicalities) that closed geodesics exist for non-simply connected manifolds, so the first non-trivial instance of the problem is the case of the 2-sphere. This was resolved by Birkhoff in 1917 [1]. Lusternic-Fet [16] generalized this to all closed manifolds in 1951.

One may ask the finer question of a bound of the length of such a closed geodesic in terms of the volume of the manifold. This is interesting even for non-simply connected Riemannian manifolds. In this case the shortest non-contractible geodesic is called the *systole* of the manifold.

The first result in this vain is due to Lowener who showed that for the 2-dimensional torus T (with any metric) one has

$$(\mathrm{sys}\, T)^2 \leq \frac{2}{\sqrt{3}}\,\mathrm{area}\, T.$$

The constant in Lowener's theorem is optimal. It is natural to ask whether one can extend Lowener's result to other surfaces or higher dimensional manifolds. Accola (1960) and Blatter (1962) gave some weak bounds for surfaces using Complex Analysis, and their bounds were improved further by Hebda and Burago in 1980. Berger, motivated by conversations with Thom, popularized the question for general manifolds in the 60's and eventually Gromov in 1983 [8] proved two quite general results. He showed for surfaces of genus g, S_g, that

$$\frac{\mathrm{sys}(S_g)^2}{\mathrm{area}\, S_g} \text{tends to } 0 \text{ as } g \to \infty \text{ and that}$$

$$\mathrm{sys}\, M^n \leq c_n \sqrt[n]{\mathrm{vol}(M^n)}$$

where c_n is a constant that depends only on the dimension of the manifold and M^n is assumed to be aspherical. We recall that M^n is called aspherical if any continuous map $f : S^k \to M^n$ extends continuously to the ball B^{k+1} (where S^k is the k-sphere, and $k \geq 2$).

Note that the n-th root is to be expected in any bound of this form as when one rescales the metric by λ the volume is multiplied by λ^n. The value of the constant c_n is significant and Gromov gave the bound

$$c_n \leq 6(n+1)n^n \sqrt{(n+1)!}.$$

Guth in 2010 improved this bound in the case of the n-dimensional torus T^n to $8n$.

We remark that there is no lower bound for the systole as one can always 'pinch' the metric making the systole arbitrarily small.

One can not expect a similar result for manifolds that are not aspherical. For example consider $S^2 \times S^1$ equipped with a Riemannian metric such that S^2 is very small and S^1 is very large, then one can arrange that the volume is 1 while the systole is arbitrarily large.

Gromov's proof of the systolic inequality was quite indirect. He used the isometric embedding $M^n \to L^\infty(M^n)$ given by $x \to \operatorname{dist}(x, \cdot)$ to define the filling radius of M^n. Here we see M^n as an n-cycle in $L^\infty(M^n)$ and the filling radius is the smallest R such that M^n bounds an $n+1$-chain in its R-neighborhood. The result follows now from two inequalities:

$$\operatorname{sys}(M^n) \leq 6 \operatorname{FillRad}(M^n)$$

$$\operatorname{FillRad}(M^n) \leq c_n \sqrt[n]{\operatorname{vol}(M^n)}.$$

The second inequality required a quite technical extension of classical isoperimetric inequalities to the infinite dimensional space $L^\infty(M^n)$.

We will outline here a more direct proof of Gromov's systolic inequality relying on [18]. This new method led to an improvement of the constant c_n by Nabutovsky to $c_n = n$ [17].

2 Uryson width and Guth's conjecture

The Uryson width is a notion of topological dimension theory that was introduced to Riemannian Geometry by Gromov [8], [9], [10].

Intuitively small k-Uryson width means that an n-dimensional space 'is close' to a k-dimensional space (where we assume $k < n$). For example if we consider a torus $T^2 = S^1 \times S^1$ where one of the S^1's has very small length ϵ and the other has, say, length 1 then T^2 is 'close' to the circle of length 1-a lower dimensional manifold.

We recall now the precise definition: if X is a metric space we say that X has q-Uryson width $\leq W$ if there exists a q-dimensional simplicial complex Y and a continuous map $\pi : X \to Y$ such that every fiber $\pi^{-1}(y)$ has diameter $\leq W$. We write then that $UW_q(X) \leq W$.

Guth ([4],[5]) proved the following theorem answering a conjecture of Gromov:

Theorem 2.1. *There exists $\epsilon_n > 0$ so that the following holds. If (M^n, g) is a closed Riemannian manifold and there exists a radius R such that every ball of radius R in (M^n, g) has volume at most $\epsilon_n R^n$ then $UW_{n-1}(M^n, g) \leq R$.*

Karasev gave recently [17] an elementary proof of the following:

Theorem 2.2. *If $(M^n; g)$ is a closed aspherical Riemannian manifold, then*

$$\mathrm{sys}(M^n) \leq 4UW_{n-1}(M^n, g).$$

Gromov had shown earlier a similar inequality but his proof used the filling radius and the inequality was weaker.

Guth's result implies Gromov's systolic inequality:

We choose R so that $\epsilon_n R^n = \mathrm{vol}(M^n)$. Then

$$\mathrm{sys}(M^n) \leq 4UW_{n-1}(M^n, g) \leq 4R = \frac{4}{\sqrt[n]{\epsilon_n}} \sqrt[n]{\mathrm{vol}(M^n)}$$

However Guth's proof is also quite technical relying on an embedding of M^n in a high dimensional cube complex, so it does not lead to an improvement of the constant in the systolic inequality.

Guth formulated a conjecture which would generalize his theorem to metric spaces. We explain this now.

In order to obtain a bound for the Uryson width of a general metric space, similar to the above result, one needs a notion of volume. One could use the n-dimensional Hausdorff measure HM_n (which coincides with Riemannian volume in the case of manifolds M^n) however this would not work well for general metric spaces. Moreover the results obtained would be far from optimal. Indeed consider an n-dimensional ϵ-thickening of an interval (with small ϵ). This metric space has a very small 1-Uryson width but its $n-1$-Hausdorff measure is infinite. It turns out that the Hausdorff content is more appropriate:

Definition 2.1. The *n-dimensional Hausdorff content* $HC_n(U)$ of a subset of a metric space X is the infimum of $\sum_{i=1}^{\infty} r_i^n$ over all coverings of U by countably many balls $B(x_i, r_i)$.

The Hausdorff content is not a measure, however it provides us with a notion of volume that is well adopted to Hausdorff width. Moreover clearly $HC_n(U) \leq HM_n(U)$ so a bound on HC_n leads to stronger results.

Conjecture (Guth)**.** *There exists $\epsilon_n > 0$ so that the following holds. If X is a metric space and there exists a radius R such that every ball B of radius R in X satisfies $HC_n(B) \leq \epsilon_n R^n$ then $UW_{n-1}(X) \leq R$.*

Guth's conjecture was proved by Liokumovich-Lishak-Nabutovsky-Rotman [15] using a method similar to the one used by Guth.

Our aim here is to outline a direct proof relying only on the co-area inequality.

We note also that Nabutovsky [17] used this method to show that one can take $c_n = n$ in Gromov's systolic inequality.

3 Proof of Guth's conjecture

We give now a sketch of the proof of Guth's conjecture before going into the details.

Let X be a compact metric space. We argue by induction. The theorem is easy to see for $n = 1$ (see Lemma 3.2 below).

The main idea in order to reduce to the lower dimension case, is to consider a subset $Y \subseteq X$ of minimal $n-1$-Hausdorff content separating the space in 'small' pieces. One can show that such a subset has locally small $n-1$-Hausdorff content by applying the co-area inequality.

By induction there is a map $f : Y \to \Sigma$ where Σ is an $n-1$-simplicial complex such that the preimages $f^{-1}(x)$ have small diameter. By adding 'cones' appropriately to Σ and using the fact that simplicial complexes are ANR's we may extend this map to $\bar{f} : X \to \Sigma'$ and obtain a map that satisfies the conditions of the conjecture.

One interesting feature of this proof is that even in the manifold case in order to carry out the induction one needs to prove the result for general metric spaces (in this case one could restrict to Riemannian polyhedra). So considering the more general context of metric spaces as suggested by Guth turns out to simplify the proof in the manifold case.

There are some technicalities to deal with: the Hausdorff content is not a measure so we work with a slight variation of this, and there is no guarantee that there is a $Y \subseteq X$ as above with minimal $n-1$-content. However, as we work with inequalities, it suffices to consider such Y which is 'nearly' minimal.

We give a definition that will allow us to sidestep the problem that Hausdorff content is not a measure so it is not additive:

Definition 3.1. The *ζ-restricted n-dimensional Hausdorff content* $HC_n^{\zeta}(U)$ of a subset of a metric space X is the infimum of $\sum_{i=1}^{\infty} r_i^n$ over all coverings of U by countably many balls $B(x_i, r_i)$ where $r_i \leq \zeta$ for all i.

Clearly we have $HC_n^{\zeta}(U) \geq HC_n(U)$. We remark that if U is contained in a ball of radius ζ then $HC_n^{\zeta}(U) = HC_n(U)$.

Notation. We denote by $B(x, r)$ the open metric ball of radius r and center x and by $\bar{B}(x, r)$ the closed ball. When we don't care about the center we denote it by $B(r)$ ($\bar{B}(r)$ respectively). We denote by $S(x, r)$ the sphere of radius r and center x, and we denote this by S_r when the center is obvious. Finally we denote by $B(r_2) \setminus B(r_1)$ the annulus between two concentric metric balls.

The co-area formula [2, Theorem 13.4.2] is our main tool. It turns out that the co-area inequality applies to Hausdorff content ([15]). We state this here for ζ-restricted Hausdorff content.

Lemma 3.1. *Let $U \subset B(r_2) \setminus B(r_1)$ be a closed set of a proper metric space. Then*

$$\int_{r_1}^{r_2} HC_{n-1}^{\zeta}(S_r \cap U)\, dr \leq 2HC_n^{\zeta}(U)$$

The same inequality applies to the Hausdorff content.

Proof. We outline a proof of this. If $B(R)$ is a ball and S_r is a sphere then $S_r \cap B(R)$ is contained in a ball of radius $\leq R$ for any r, so $HC_{n-1}(S_r \cap B(R)) \leq R^{n-1}$ for any r. So if $B(R)$ is a ball contained in an annulus $B(r_2) \setminus B(r_1)$ and $\zeta \geq R$ we have

$$\int_{r_1}^{r_2} HC_{n-1}^{\zeta}(S_r \cap B(R))\, dr \leq 2R \cdot R^{n-1} \quad (*).$$

Note now that if U is any closed set for any $\epsilon > 0$ there is a covering of U by finitely many balls $B_i(r_i)$, $i = 1, ..., k$ so that $r_i \leq \zeta$ and $\sum_{i=1}^k r_i^n - HC_n^{\zeta}(U) < \epsilon$ and the result follows by $(*)$. Clearly this proof applies to $HC_n(U)$ as well. □

We treat now the case $n = 1$.

Lemma 3.2. *Let X be a proper metric space and let $R > 0$. If for any $x \in X$ the 1-dimensional Hausdorff content of the ball $B(x, R)$ is bounded by $\dfrac{1}{100}R$ then $UW_0(X) \leq R$.*

Proof. We set $\delta = \dfrac{1}{100}R$. We fix $x_0 \in X$ and we consider the closed annuli $A_k = \{x \in X : 10(k-1)R \leq d(x_0, x) \leq 10kR\}$, $k \geq 1, k \in \mathbb{N}$. Each A_k is compact so it has a finite covering by balls $B_j(r_j)$ such that $r_j \leq 2\delta$ for all j. Let

$$a_k = HC_1^{2\delta}(A_k).$$

We pick for each A_k a covering by open balls $B_j(r_j)$ such that

$$\sum r_j - a_k < \delta \quad (*).$$

By doing this for all k we obtain a covering $\mathcal{U}$ of X by open balls.

Suppose that we have a finite sequence of balls in $\mathcal{U}$, $B_1(r_1), ..., B_n(r_n)$ such that $B_i(r_i)$ intersects $B_{i+1}(r_{i+1})$ for all i. We claim that if this happens then

$$\sum_{i=1}^{n} r_i \leq 10\delta.$$

We may assume by taking a smaller n if necessary and arguing by contradiction that

$$12\delta \geq \sum_{i=1}^{n} r_i > 10\delta.$$

So all these balls are contained in a ball $B(x, R)$ which is contained either in a single annulus A_k or in a union of two annuli $A_k \cup A_{k+1}$. However by our hypothesis the content of $B(x, R)$ is bounded by δ, so we could replace these balls in $\mathcal{U}$ by finitely many balls $B_s(r_s)$, $s \in S$ such that their union contains $B(x, R)$ and

$$\sum_{s \in S} r_s < 2\delta.$$

It follows that the sequence $B_1(r_1), ..., B_n(r_n)$ violates $(*)$ for at least one of A_k, A_{k+1}.

Let $B \in \mathcal{U}$. We note now that if $B_1(r_1), ..., B_n(r_n)$ is a finite sequence of balls from $\mathcal{U}$ containing B such that $B_i(r_i)$ intersects $B_{i+1}(r_{i+1})$ their union has diameter $< R/2$.

We introduce an equivalence relation on $\mathcal{U}$. We say that two balls B, B' in $\mathcal{U}$ are equivalent if there is a finite sequence of balls $B_1 = B, B_2, ..., B_n = B'$ such that any two successive balls in the sequence intersect.

We replace then each equivalence class of balls from $\mathcal{U}$ by their union.

In this way we obtain a cover of X by sets say D_i, $i \in \mathbb{N}$ such that each D_i is open (as a finite union of open balls), and closed (since its complement is open). It follows that the map $f : X \to \mathbb{N}$ where $f(D_k) = k$ is continuous and

$$\mathrm{diam} f^{-1}(k) = \mathrm{diam} D_k < R$$

so $UW_0(X) \leq R$.

$\square$

If U is an open subset of a Riemannian manifold then $\mathrm{vol}_n(U)$ is equal to the n-Hausdorff measure of U which is in turn greater or equal to the n-dimensional Hausdorff content.

Theorem 3.3. *There is an $\epsilon_n > 0$ such that the following holds. If X is a compact metric space such that for any $x \in X$ the n-dimensional Hausdorff content of the ball $B(x, R)$ is bounded by $\epsilon_n R^n$ then $UW_{n-1}(X) \leq R$.*

Proof. We will prove by induction on n that there is a continuous map $\pi : X \to \Sigma$ where Σ is a finite simplicial complex of dimension $\leq n-1$ such that diam $\pi^{-1}(y) \leq R$ for any $y \in \Sigma$. The theorem holds for $n = 1$ by Lemma 3.2.

Definition 3.2. Let $Z \subseteq X$ closed. We say that Z is a *D-separating subset* if

$$X \setminus Z = \bigsqcup_{i \in I} U_i$$

where the U_i are open disjoint sets of diameter $\leq D$ and I is finite. We say that the open sets U_i are the *pieces of the decomposition of X by Z*.

We set $\zeta = R/1000$. Let $b(D)$ be the infimum of $HC^{\zeta}_{n-1}(Z)$ over all D-separating sets Z. It is not clear whether there exists a D-separating set realizing $b(D)$ however it will be sufficient for us to consider sets with content close enough to $b(D)$: We say that Z is a *δ-minimal D-separating set* if Z is D-separating and

$$HC^{\zeta}_{n-1}(Z) - b(D) \leq \delta.$$

In what follows our statements will be true for δ sufficiently small.

The theorem follows from the next lemma:

Lemma 3.4. *There is an $\epsilon_n > 0$ such that the following holds. If X is a compact metric space such that for any $x \in X$ the n-dimensional Hausdorff content of the ball $B(x, R)$ is bounded by $\epsilon_n R^n$ then there is a finite simplicial complex Σ of dimension $\leq n-1$ and a continuous map $f : X \to \Sigma$ such that: diam $f^{-1}(e) \leq R$ for any simplex $e \in \Sigma$.*

Proof. We prove this by induction on n. For $n = 1$ the statement follows by Lemma 3.2. In particular we may take $\epsilon_1 = 1/100$.

We will show that the lemma holds for ϵ_n where we define ϵ_n inductively by $\epsilon_n = \epsilon_{n-1}/1000^{n+1}$.

We assume now that the lemma holds for $n-1$ for some $n \geq 2$.

Lemma 3.5. *Let ϵ_{n-1} be the constant provided by Lemma 3.4 and let $\epsilon_n = \epsilon_{n-1}/1000^{n+1}$. Let X be a compact metric space such that for any $x \in X$ the n-dimensional Hausdorff content of the ball $B(x, R)$ is bounded by $\epsilon_n R^n$. Let Z be a δ-minimal $R/4$-separating subset of X. Then for any ball of radius $R/1000$, $B(x, R/1000)$,*

$$HC^{\zeta}_{n-1}(Z \cap B(x, R/1000)) \leq \epsilon_{n-1}\big(\frac{R}{1000}\big)^{n-1}.$$

Proof. We argue by contradiction assuming that Z does not satisfy this inequality for some x. We take

We note that $(R/1000)^n \geq \epsilon_n R^n$. It follows that $HC_n(B(x, R)) = HC^{\zeta}_n(B(x, R))$. By the co-area inequality (Lemma 3.1) and our hypothesis that $HC^{\zeta}_n(B(x, R)) \leq \epsilon_n R^n$ we have that for some $r \in [R/100, R/50]$

$$HC^{\zeta}_{n-1}S(x, r) \leq 200\epsilon_n R^{n-1} \leq \frac{\epsilon_{n-1}R^{n-1}}{5 \cdot 1000^n}.$$

If $Z_1 = S(x, r)$ and $Z_2 = B(x, r) \cap Z$ we set $Z' = (Z \setminus Z_2) \cup Z_1$. We claim that Z' is $R/4$-separating. Indeed let

$$X \setminus Z = \bigsqcup_{i \in I} U_i$$

where I is finite and the U_i are open disjoint sets of diameter $\leq R/4$. Let $U = B(x, r)$. Then

$$X \setminus Z' = \bigsqcup_{i \in I} (U_i \setminus \bar{B}(x, r)) \sqcup U.$$

If $B_i(r_i), i \in I$ is a cover of Z by balls of radius $\leq \zeta$ so that

$$\sum_{i \in I} r_i^{n-1} - HC^{\zeta}_{n-1}(Z) < \delta$$

we get a cover of Z' by omitting all balls from this cover intersecting $B(x, R/1000)$ and adding appropriately balls that cover $S(x, r)$ and approximate $HC^{\zeta}_{n-1}S(x, r)$ up to δ.

We have then

$$HC^{\zeta}_{n-1}(Z') \leq HC^{\zeta}_{n-1}(Z) - \epsilon_{n-1}\big(\frac{R}{1000}\big)^{n-1} + \frac{\epsilon_{n-1}R^{n-1}}{5 \cdot 1000^n} + \delta$$

contradicting the δ-minimality property of Z if we take

$$\delta < \frac{\epsilon_{n-1}R^{n-1}}{1000^n}.$$

□

We prove now Lemma 3.4. Let Z be a δ-minimal $R/4$-separating subset of X. By Lemma 3.5 and our inductive hypothesis there is a continuous map $\pi_1 : Z \to \Sigma_1$ where Σ_1 is a finite simplicial complex of dimension $\leq n-2$ such that $\operatorname{diam} \pi_1^{-1}(e) \leq R/1000$ for any simplex $e \in \Sigma_1$.

Let U be a piece of the decomposition of X by Z. Clearly $\partial U \subset Z$ so $\pi_1(\partial U)$ is contained in a finite subcomplex of Σ_1. We denote by Σ_U the minimal such subcomplex of Σ_1.

We define a new simplicial complex Σ as follows: For each closure of a connected component U we consider the cone C_U over Σ_U (which is a simplicial complex of dimension $\leq n-1$). We glue C_U to Σ_1 along their common subcomplex Σ_U.

We will need some facts from topology that we recall now (see eg [12]). Any finite simplicial complex is an Absolute Neighborhood Retract (ANR). A contractible ANR is an Absolute Retract (AR). In particular the cone of a finite simplicial complex is an AR. A space A is an AR if and only if it is an absolute extensor i.e. if it has the following property: if B is any metric space, $K \subseteq B$ is closed and $f : K \to A$ is continuous then f can be extended continuously to the whole of B.

By the above facts it follows that for each U the map $\pi_1 : \partial U \to \Sigma_U \subset C_U$ can be extended to a continuous map $\pi : U \to C_U \subset \Sigma$. Since X is the union of Z with the pieces of the decomposition of X by Z and since the map π is continuous on the closure of each piece we have that the map $\pi : X \to \Sigma$ is continuous.

Let e be a maximal simplex of Σ. Then e is either a simplex of Σ_1 or a cone of a simplex e' of Σ_1. If $\pi(U)$ intersects e then in the first case ∂U intersects $\pi_1^{-1}(e)$ while in the second case ∂U intersects $\pi_1^{-1}(e')$. Since

$$\operatorname{diam} \pi_1^{-1}(e') \leq R/1000 \text{ and } \operatorname{diam}(U) \leq R/4$$

we have that

$$\operatorname{diam} \pi^{-1}(e) \leq R.$$

□

Clearly the theorem follows from the lemma as any point of Σ is contained in some simplex e of Σ.

□

As we remarked in the previous section this Theorem gives a new proof of the systolic inequality for aspherical manifolds. It turns out that the result is also valid for proper metric spaces.

4 Final remarks and open problems

Gromov's work on systolic inequalities gave rise to a branch of geometry called often systolic geometry or curvature free geometry. This is because one tries to find relations between the geometry and topology of the manifold that apply to all manifolds without any curvature restrictions. For a beautiful introduction to the subject we refer to Guth's

ICM talk [7]. There is a number of interesting open problems and we state some here to give a flavor of the subject:

Is there a bound for the length of the shortest periodic geodesic in terms of volume for general manifolds? Croke, answering a conjecture of Gromov, showed that this is the case for the 2-sphere S^2 but this is not known for the 3-sphere S^3.

The geometry a surfaces of high genus is not very well understood. A major unresolved question is Buser's conjecture, stating that there is a constant c such that any surface S_g admits a pants decomposition where all curves of the decomposition have length bounded by $c\sqrt{\operatorname{area} S_g}$. The necessity of length at least $\sim \sqrt{\operatorname{area} S_g}$ can be seen by considering a standard sphere with many small handles attached. In fact one may think of such surfaces as a 2-dimensional analog of expander graphs. However unlike the case of expanders we don't have good ways to construct random surfaces.

Is it true that among all orientable surfaces of area 1 the torus has the largest systole? This is known for surfaces of genus $g = 2$ and $g \geq 20$.

References

[1] Birkhoff, G. D. *Dynamical systems with two degrees of freedom.* Proceedings of the National Academy of Sciences of the United States of America, 3(4), 314, (1917).

[2] Burago, Y. D. and Zalgaller, V. A. *Geometric inequalities*, Vol. 285. Springer Science & Business Media, 2013.

[3] Federer, H. *Geometric measure theory.* Springer, 2014.

[4] Guth, L. *Volumes of balls in large Riemannian manifolds.* Annals of mathematics 173.1 (2011): 51-76.

[5] Guth, L. *Volumes of balls in Riemannian manifolds and Uryson width.* Journal of Topology and Analysis 9.02 (2017): 195-219.

[6] Guth, L. *Systolic inequalities and minimal hypersurfaces.* Geometric and Functional Analysis. 2010 Mar 1;19(6):1688-92.

[7] Guth, L. *Metaphors in systolic geometry.* Proceedings of the International Congress of Mathematicians 2010 (ICM 2010) (In 4 Volumes) Vol. I: Plenary Lectures and Ceremonies Vols. II-IV: Invited Lectures. 2010.

[8] Gromov, M. *Filling riemannian manifolds* , Journal of Differential Geometry 18.1 (1983): 1-147.

[9] Gromov, M. *Width and related invariants of Riemannian manifolds.* Asterisque 163.164 (1988): 93-109.

[10] Gromov, M. *Large Riemannian manifolds.* Curvature and topology of Riemannian manifolds. Springer, Berlin, Heidelberg, 1986. 108-121.

[11] Gromov, M. *Volume and bounded cohomology*, Inst. Hautes Etudes Sci. Publ. Math. (1982), 56:5-99 (1983).

[12] Hu, S. T. *Theory of retracts.* Wayne State University Press (1965).

[13] Hardt, R., and Simon L. *Seminar on geometrical measure theory* Birkhauser Basel, 1986.

[14] Hurewicz, W. and Wallman, H. *Dimension theory* (Vol. 4). Princeton university press, 1969.

[15] Liokumovich, Y., Lishak, B., Nabutovsky, A., Rotman, R. *Filling metric spaces* . (2019). arXiv preprint arXiv:1905.06522

[16] Lyusternik, L. A. and Fet A. I. *Variational problems on closed manifolds*, Dokl. Akad. SSSR (N. S.) 81 (1951), 17-18 (Russian).

[17] Nabutovsky, A. *Linear bounds for constants in Gromov's systolic inequality and related results* . (2019). arXiv preprint arXiv:1909.12225

[18] Papasoglu, P. *Uryson width and volume.* Geometric and Functional Analysis, 30(2), 574-587, (2020).

[19] Willard S. *General topology.* Courier Corporation; 2004.

Panos Papasoglu: Mathematical Institute, University of Oxford, Andrew Wiles Building, Woodstock Rd, Oxford OX2 6GG, U.K.

E-mail: papazoglou@maths.ox.ac.uk