REPORTS@SCM

AN ELECTRONIC JOURNAL OF THE SOCIETAT CATALANA DE MATEMÀTIQUES

Volume 10, num. 1, 2025

ISSN (electronic edition): 2385-4227

https://revistes.iec.cat/index.php/reports

Institut d'Estudis Catalans

REPORTS@SCM

AN ELECTRONIC JOURNAL OF THE
SOCIETAT CATALANA DE MATEMÀTIQUES

Volume 10, number 1

December 2025

https://revistes.iec.cat/index.php/reports
ISSN electronic edition: 2385-4227

Societat Catalana de Matemàtiques

Institut d'Estudis Catalans

Focus and Scope

Reports@SCM is a non-profit electronic research journal on Mathematics published by the Societat Catalana de Matemàtiques (SCM) which originated from the desire to help students and young researchers in their first steps into the world of research publication.

Reports@SCM publishes short papers in all areas of pure mathematics, applied mathematics, and mathematical statistics, including also mathematical physics, theoretical computer science, and any application to science or technology where mathematics plays a central role. To be considered for publication in *Reports@SCM* an article must be written in English (with an abstract in Catalan), be mathematically correct, and contain some original interesting contribution. All submissions will follow a peer review process before being accepted for publication.

Research announcements containing preliminary results of a larger project are also welcome. In this case, authors are free to publish in the future any extended version of the paper elsewhere, with the only condition of making an appropriate citation to *Reports@SCM*.

We especially welcome contributions from researchers at the initial period of their academic careers, such as Master or PhD students. We wish to give special attention to the authors during the whole editorial process. We shall take special care to maintain a reasonably short average time between the reception of a paper and its acceptance, and between its acceptance and its publication.

All manuscripts submitted will be **peer reviewed** by at least one reviewer. Final decisions on the acceptance of manuscripts are taken by the editorial board, based on the reviewer's opinion.

Edited by Societat Catalana de Matemàtiques, Institut d'Estudis Catalans (IEC)
Carrer del Carme 47, 08001 Barcelona.
https://scm.iec.cat Telèfon: (+34) 93 324 85 83
scm@iec.cat Fax: (+34) 93 270 11 80

Institut d'Estudis Catalans
https://www.iec.cat
informacio@iec.cat

https://revistes.iec.cat/index.php/reports
ISSN electronic edition: 2385-4227

Reports@SCM
Volume 10, number 1, 2025

Table of Contents

REPORTS@SCM
AN ELECTRONIC JOURNAL OF THE
SOCIETAT CATALANA DE MATEMÀTIQUES

Generating uniform spanning trees from conditioned Bienaymé–Galton–Watson trees

Albert Vives
Universitat Politècnica de Catalunya
avivesbatalla@gmail.com

Resum *(CAT)*
Aquest article explora la generació d'arbres generadors uniformes (UST), fonamentals en combinatòria i probabilitat, amb aplicacions en teoria de xarxes i física. Utilitzant processos de Bienaymé–Galton–Watson (BGW) condicionats a un nombre fix de vèrtexs, s'introdueix un mètode per generar arbres generadors uniformes i s'examinen propietats estructurals com l'alçada i l'amplada.

Abstract *(ENG)*
This report explores uniform spanning tree (UST) generation, essential in combinatorics and probability with applications in network theory and physics. Using conditioned Bienaymé–Galton–Watson (BGW) processes, it introduces a method to generate USTs. Rigorous proofs show that conditioning on a fixed number of vertices ensures uniform distribution and let us examine structural properties like height and width.

Societat Catalana de Matemàtiques
Institut d'Estudis Catalans

Keywords: *random trees, Bienaymé–Galton–Watson trees.*
MSC (2020): *Primary 05C05. Secondary 60J80.*

Received: *February 17, 2025.*
Accepted: *October 2, 2025.*

1. Introduction

The generation of uniform spanning trees (USTs) is a crucial problem in combinatorics and stochastic processes, with applications in network analysis, random structures, and statistical physics. A spanning tree is a connected subgraph of a graph that includes all vertices without forming cycles. When sampled uniformly at random, each spanning tree has an equal probability of being chosen, presenting mathematical challenges in defining an appropriate random model susceptible of being analyzed. This drives the search for effective approaches, and this report proposes novel methods using conditioned Bienaymé–Galton–Watson (BGW) processes. These processes, traditionally used to model population growth, are adapted to generate spanning trees under uniform distributions, bridging combinatorics with stochastic processes.

This work provides detailed mathematical formulations and rigorous proofs demonstrating that conditioning BGW processes on a fixed number of vertices results in uniform spanning trees. It explores classical distributions such as Poisson, Geometric, and Bernoulli branching to illustrate the method's versatility and correctness. It also delves into structural characteristics of random trees, analyzing height and width as primary metrics for complexity. Results on these parameters offer insights into the asymptotic behavior of USTs, supported by probabilistic bounds and examples. These findings contribute to a deeper understanding of USTs and their generation through probabilistic techniques.

2. Obtaining uniform spanning trees from BGW trees

Depending on the probability distribution governing the reproduction of some individuals, one may obtain different types of trees. Specifically, we aim to prove that if we condition a Bienaymé–Galton–Watson tree to have n vertices and fix certain known distributions, we obtain various types of uniform random trees. The first description of random trees from conditioning a BGW process by its total progeny can be traced back to Kolchin [3] and Aldous [2].

Lemma 2.1. *Let Z be a nonnegative integer-valued random variable and let X be a* BGW(Z) *process. Let $\mathcal{T}_n$ denote the class of trees with n vertices which can be generated by the process and let $T \in \mathcal{T}_n$ be a tree in the class. If the probability that T is generated by the process depends only on the number of vertices n, then*

$$\mathbb{P}(X = T||X| = n) = \frac{1}{|\mathcal{T}_n|}.$$

Proof. Let us denote $\mathbb{P}(X = T)$ by $f(n)$, where n is the number of vertices of T. Then,

$$\mathbb{P}(|X| = n) = |\mathcal{T}_n| f(n),$$

since the last probability is the sum of the probabilities of obtaining each of the trees with n vertices. Now, when conditioned on having n vertices, all possible trees are equally likely to occur:

$$\mathbb{P}(X = T \mid |X| = n) = \frac{\mathbb{P}(\{X = T\} \cap \{|X| = n\})}{\mathbb{P}(|X| = n)} = \frac{f(n)}{|\mathcal{T}_n| f(n)} = \frac{1}{|\mathcal{T}_n|}.$$ □

We now aim to prove that, by selecting an appropriate offspring distribution, it is possible to generate several well-known types of trees.

The class of labeled trees with n nodes is also called the class of Cayley trees, due to the Cayley formula enumerating them, its number being n^{n-2}. While Bienaymé–Galton–Watson trees are naturally rooted, Cayley trees are not; in this context we consider Cayley trees to be rooted by fixing one distinguished vertex as the root.

Theorem 2.2. *Conditioning a Bienaymé–Galton–Watson tree with offspring distribution* Poisson(1) *on having n vertices results in a Cayley tree with n vertices generated uniformly at random.*

Proof. Every rooted Cayley tree can be generated from a Poisson(1) BGW process by a labeling of its vertices. Let T be a specific Cayley tree with n vertices. To prove that each tree can be obtained uniformly, we first need to order all sibling sets in T by increasing vertex labels. Let $\chi_1, \dots, \chi_n$ represent the number of children of each node, where the vertices are indexed starting from the root and then recursively visiting the children from left to right. The first requirement for generating T is ensuring the correct number of descendants for each vertex. Since these random variables are mutually independent, the probability of obtaining a specific number of children for all vertices is the product of their individual probabilities.

The second requirement is assigning the correct labeling, as we are considering labeled trees. With n vertices, there are $n!$ possible ways to label them. Moreover, the children of the i-th vertex can be permuted in $\chi_i!$ distinct ways for each $i = 1, \dots, n$, resulting in the same tree. Thus, the final calculation can be expressed clearly as follows:

$$\mathbb{P}(X = T) = \left(\prod_{i=1}^{n} \frac{1}{\chi_i!} \cdot e^{-1} \right) \cdot \frac{\prod_{i=1}^{n} \chi_i!}{n!} = e^{-n} \frac{1}{n!}.$$

Since the last probability depends only on the number of vertices, and it is known that there exist n^{n-2} labeled trees, by Lemma 2.1 the theorem is proved. □

The class $\mathcal{B}_n$ of full binary trees of n vertices is the family of unlabelled rooted plane trees where every node has two or zero children. Being plane means that trees have distinguished left and right subtrees. The trees depicted in Figure 1 are considered to be distinct.

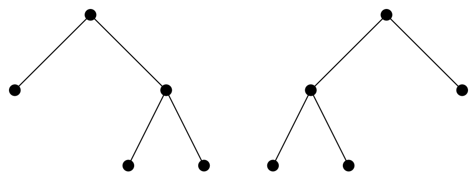

Figure 1: Two distinct binary trees.

A full binary tree with n nodes has an odd number n of vertices and $m = (n+1)/2$ leaves. The number of such trees is given by the Catalan number C_{m-1}.

Theorem 2.3. *Conditioning a* 2 Bernoulli(1/2) *Bienaymé–Galton–Watson tree on having n vertices results in a binary tree with n vertices generated uniformly at random.*

Proof. It is clear that one can obtain every binary tree from a $2\,\mathrm{Be}(1/2)$ BGW tree and vice versa. Let T be a particular binary tree with n vertices. It is known that in a binary tree there are $(n-1)/2$ internal nodes, and $(n+1)/2$ leafs. To apply Lemma 2.1, we need to calculate $\mathbb{P}(X = T)$. Let $\chi \sim 2\,\mathrm{Be}(1/2)$.

$$\mathbb{P}(X = T) = \mathbb{P}(\chi = 0)^{\frac{n+1}{2}}\mathbb{P}(\chi = 2)^{\frac{n-1}{2}} = \left(\frac{1}{2}\right)^n .$$

Since the last probability depends just on the number of vertices, by Lemma 2.1 we have proved the theorem. □

The class $\mathcal{P}_n$ of ordered plane trees with n vertices is the family of unlabelled rooted trees where the children of every node are ordered from left to right in the plane.

Theorem 2.4. *Conditioning a* Geom$(1/2)$ *Bienaymé–Galton–Watson tree on having n vertices results in an ordered plane tree with n vertices generated uniformly at random, where* Geom$(1/2)$ *denotes the geometric distribution on* $\{0, 1, 2, \dots\}$.

Proof. It can be observed that every ordered plane tree can be obtained from a Geom$(1/2)$ BGW tree, and conversely, every Geom$(1/2)$ BGW tree corresponds to an ordered plane tree. Let T be a specific ordered plane tree with n vertices. To generate such a tree, it is necessary to account for both internal nodes and leaves.

The probability of an internal node having exactly k children is $(1/2)^k(1/2)$, where the first factor represents the probability of successfully having k children, and the second factor accounts for the probability of no additional children. For a leaf, the requirement is simply to have no children, which occurs with probability $1/2$.

Consequently, the probability of achieving the correct number of children for each vertex is independent of whether the vertex is an internal node or a leaf. This probability is given by $(1/2)^{\chi_i+1}$, where χ_i denotes the number of children of vertex i. Thus, the following conclusion naturally arises:

$$\mathbb{P}(\mathcal{T} = T) = \prod_{i=1}^{n}\left(\frac{1}{2}\right)^{\chi_i+1} = \left(\frac{1}{2}\right)^{\sum_{i=1}^{n}\chi_i+1} = \left(\frac{1}{2}\right)^{n-1+n} = \left(\frac{1}{2}\right)^{2n-1}.$$

Furthermore, it is usually known that the number of ordered plane trees with n vertices is C_{n-1}, where C_n is the n-th Catalan number. Hence, by Lemma 2.1, the theorem has been proved. □

3. Studying some parameters of random trees

This chapter is focused on the study of key parameters of random trees: height, width, and the number of leaves. All results presented in this section are derived from the study L. Addario-Berry, L. Devroye, and S. Janson did in [1]. The approach in [1] applies not only to random Cayley trees, but to any family of trees arising from a BGW(χ) tree as long χ has expectation 1 and finite variance (critical BGW trees). Furthermore, explicit proofs are provided for certain concepts that are often assumed without further justification in the literature. The exploration of these parameters offers a deeper understanding of random trees, contributing new insights into established results in this field.

3.1. Some preliminaries

The Breadth-First Search (BFS) on a BGW tree is an algorithm used to explore the tree level by level, starting from the root. BFS explores the tree by visiting all nodes at one level before moving to the next one, ensuring that nodes are processed in increasing distance from the root. Hence, this search keeps a queue Q with Q_i nodes at the i-th step, with $Q_0 = 1$. During the exploration of a vertex, its offspring are added to the back of the queue. Then, one can easily obtain the following recursion:

$$Q_i = Q_{i-1} - 1 + \chi_i,$$

where χ_i are independent and identically distributed copies of the offspring distribution χ. Hence, by this recursion, $Q_j = 1 + \tilde{S}_j$, where $\tilde{S}_j := \sum_{i=1}^{j}(\chi_i - 1) = S_j - j$. The tree is completely explored when $Q_j = 0$. In this case, $\tilde{S}_n = -1$.

Definition 3.1. Let T_n be a random tree with n nodes. The width of a tree T_n, denoted by $W(T_n)$, is the maximum number of nodes at any depth level. Generally, let $d(v)$ denote the depth of a node v in a rooted tree T. Then

$$\text{width}(T_n) = \max_{k \geq 0} |\{v \in V(T_n) : d(v) = k\}|.$$

We now present key lemmas that are necessary for studying the expected width of a random tree.

Lemma 3.2 (Raney's Lemma)**.** *Let $a_1, a_2, \dots, a_n$ be a sequence of integers such that $\sum_{i=1}^{n} a_i = -1$.*

Then there exists a unique index s such that the cyclic partial sums $S_k = \sum_{j=0}^{k-1} a_{(s+j) \bmod n}$ for $k = 1, 2, \dots, n$, satisfy:

(i) $S_k > 0$ *for* $1 \leq k < n$.

(ii) $S_n = -1$.

Lemma 3.3. *Suppose that the individuals in a BGW process reproduce according to a random variable χ, with $\mathbb{E}[\chi] = 1$ and $\text{Var}(\chi) < \infty$. Then, there is a constant $c_1 \in \mathbb{R}$ such that, for all n sufficiently large,*

$$\mathbb{P}(\tilde{S}_n = -1) \geq c_1 n^{-1/2}.$$

Lemma 3.4. *Suppose that χ_i are i.i.d., non-negative and integer-valued random variables, with $\mathbb{E}[\chi_i] = 1$ and $\text{Var}[\chi_i] < \infty$, and let $S_n = \sum_{i=1}^{n} \chi_i$. Then, for all $n \geq 1$ and $m \geq 0$,*

$$\mathbb{P}(S_n = n - m) \leq \frac{c_2}{\sqrt{n}} e^{-c_3 m^2/n},$$

where $c_2 > 0$ and $c_3 > 0$ are real constants.

Lemma 3.5. *Let χ be a discrete random variable taking values in nonnegative integers. Suppose that $\mathbb{E}(\chi) = 1$ and $0 < \text{Var}(\chi) < \infty$. Let $\mathcal{T}$ be a BGW(χ) tree. Then, $\mathbb{P}(|\mathcal{T}| = n) \geq n^{-3/2}$.*

Lemma 3.6 (Bernstein inequality)**.** *Let $X_1, X_2, \dots, X_n$ be independent random variables such that $X_i - \mathbb{E}[X_i] \leq b$ for every i, where $b \in \mathbb{R}$. Let $V := \sum_{i=1}^{n} \text{Var}(X_i)$. Then,*

$$\mathbb{P}\left(\sum_{i=1}^{n}(X_i - \mathbb{E}X_i) \geq t\right) \leq \exp\left(-\frac{t^2}{2V + \frac{2bt}{3}}\right).$$

3.2. The width

Theorem 3.7. *Let χ be a discrete random variable taking values in nonnegative integers. Suppose that $\mathbb{E}[\chi] = 1$ and $0 < \mathrm{Var}(\chi) < \infty$. Let T_n be a BGW(χ) tree with n nodes. Then,*

$$\mathbb{P}(W(T_n) \geq x) \leq c_4 e^{-c_5 x^2/n},$$

for all $x \geq 0$ and $n \geq 1$, where $c_4 > 0$ and $c_5 > 0$.

Proof. We will follow the proof strategy outlined in [1]. Let Z_k be the number of individuals in the k-th level of a BGW tree. It is clear that every Z_k is some Q_j, where Q_j is the number of nodes in the queue of BFS in the j-th iteration. Hence,

$$W := \max_{k \geq 0} Z_k \leq \max_{j \geq 0} Q_j.$$

As a result, for the conditioned BGW tree T_n,

$$\mathbb{P}(W \geq x+1) \leq \mathbb{P}(\max_j Q_j \geq x+1) = \mathbb{P}(\max_j \tilde{S}_j \geq x \mid \tilde{S}_j \geq 0,\, j < n,\, \tilde{S}_n = -1).$$

Now, we aim to simplify the last conditioned probability. Our first goal is to get rid of the conditioning on $\tilde{S}_j \geq 0$, where $j < n$. By Raney's Lemma 3.2, conditioning on $\{\tilde{S}_j \geq 0,\, j < n,\, \tilde{S}_n = -1\}$ is equivalent to conditioning only on $\{\tilde{S}_n = -1\}$. However, $\max_j \tilde{S}_j$ may be changed. By conditioning on $\tilde{S}_n = -1$, we can write

$$\max_{j \leq n} \tilde{S}_j = \max_{j \leq n} \tilde{S}_j - \min_{j \leq n} \tilde{S}_j + 1,$$

and the latter quantity is changed by at most 1 by a rotation of $\tilde{\chi}_i := \chi_i - 1\ \forall i = 1, \dots, n$. Then

$$\mathbb{P}(\max_j \tilde{S}_j \geq x \mid \tilde{S}_j \geq 0,\, j < n,\, \tilde{S}_n = -1) \leq \mathbb{P}(\max_{j \leq n} \tilde{S}_j - \min_{j \leq n} \tilde{S}_j \geq x \mid \tilde{S}_n = -1).$$

Therefore, we now have more tools to bound $\mathbb{P}(W \geq x+1)$.

$$\begin{aligned}\mathbb{P}(W \geq 2x+2) \leq \mathbb{P}(\max_j Q_j \geq 2x+2) &\leq \mathbb{P}(\max_{j \leq n} \tilde{S}_j - \min_{j \leq n} \tilde{S}_j \geq 2x+1 \mid \tilde{S}_n = -1) \\ &\leq \mathbb{P}(\max_{j \leq n} \tilde{S}_j \geq x \mid \tilde{S}_n = -1) + \mathbb{P}(\min_{j \leq n} \tilde{S}_j \leq -x-1 \mid \tilde{S}_n = -1).\end{aligned}$$

The last inequality is due to the fact that, if $y - z \geq 2x+1$, then either $y \geq x$ or $z \leq -x-1$, where x, y, z are real numbers. Furthermore, the reflection $\chi_i \leftrightarrow \chi_{n+1-i}$, which swaps $\tilde{S}_j \leftrightarrow -\tilde{S}_n - \tilde{S_{n-j}}$, shows that the last probabilities are the same. Hence,

$$\mathbb{P}(\max_j Q_j \geq 2x+2) \leq 2\,\mathbb{P}(\max_{j \leq n} \tilde{S}_j \geq x \mid \tilde{S}_n = -1).$$

The last expression can be written in terms of the first index such that $\tilde{S}_j \geq x$. So that, let us define $\tau = \min\{j \geq 0 : \tilde{S}_j \geq x\}$. Then,

$$\begin{aligned}\mathbb{P}(\max_j Q_j \geq 2x+2) &\leq 2\,\mathbb{P}(\tau < n \mid \tilde{S}_n = -1) \\ &= 2\,\mathbb{P}(\tilde{S}_n = -1 \mid \tau < n)\frac{\mathbb{P}(\tau < n)}{\mathbb{P}(\tilde{S}_n = -1)},\end{aligned}$$

where the last equality is due to the definition of conditioned probability. By definition of τ, $\tilde{S}_\tau \geq x$. Then, if we fix some $t < n$ and $y \geq x$, by Lemma 3.4 we have the following:

$$\begin{aligned}\mathbb{P}(\tilde{S}_n = -1 \mid \tau < n) &= \mathbb{P}(\tilde{S}_n = -1 \mid \tau = t, \tilde{S}_\tau = y)\\ &= \mathbb{P}(\tilde{S}_n - \tilde{S}_t = -y-1) = \mathbb{P}(\tilde{S}_{n-t} = -(y+1)) = \mathbb{P}(S_{n-t} - (n-t) = -(y+1))\\ &= \mathbb{P}(S_{n-t} = (n-t) - (y+1)) \leq c_2 n^{-1/2} e^{-c_3(y+1)^2/(n-t)}\\ &\leq c_2 n^{-1/2} e^{-c_3 x^2/n},\end{aligned} \tag{1}$$

where we have used $\tilde{S}_n - \tilde{S}_t = \tilde{S}_{n-t}$ due to the fact that these $n-t$ random variables are i.i.d. Then, by (1) and Lemma 3.3 it is clear that we can now prove what we seek:

$$\mathbb{P}(\max_j Q_j \geq 2x+2) \leq c_2 n^{-1/2} e^{-c_3 x^2/n} \cdot \frac{\mathbb{P}(\tau < n)}{\mathbb{P}(\tilde{S}_n = -1)} \leq \frac{c_5 n^{-1/2} e^{-c_7 x^2/n}}{\mathbb{P}(\tilde{S}_n = -1)} \leq c_4 e^{-c_5 x^2/n}. \qquad \square$$

Theorem 3.8. *Let χ be a discrete random variable taking values in nonnegative integers. Suppose that $\mathbb{E}[\chi] = 1$ and $0 < \mathrm{Var}(\chi) < \infty$. Let T_n be a $\mathrm{BGW}(\chi)$ tree with n nodes. Then,*

$$\mathbb{E}[W(T_n)] = \mathcal{O}(\sqrt{n}),$$

for all $n \geq 1$.

Proof. This proof is not provided in [1]; however, we consider it highly relevant to the topic at hand. Therefore, we present a rigorous proof of this result herein. We aim to prove that the expected value $\mathbb{E}[W(T_n)]$ grows asymptotically at most as $\sqrt{n}$, given the inequality

$$\mathbb{P}(W(T_n) \geq w) \leq c_4 e^{-c_5 w^2/n}.$$

For a discrete random variable W, the expectation can be written as

$$\mathbb{E}[W(T_n)] = \sum_{w=0}^{\infty} w \cdot \mathbb{P}(W(T_n) = w),$$

or equivalently

$$\mathbb{E}[W(T_n)] = \sum_{w=1}^{\infty} \mathbb{P}(W(T_n) \geq w).$$

This equivalence holds because

$$\mathbb{P}(W(T_n) \geq w) = \sum_{k=w}^{\infty} \mathbb{P}(W(T_n) = k),$$

and reorganizing terms yields the alternative representation.

Using the given inequality $\mathbb{P}(W(T_n) \geq w) \leq c_4 e^{-c_5 w^2/n}$, we can bound the expectation as

$$\mathbb{E}[W(T_n)] \leq \sum_{w=1}^{\infty} c_4 e^{-c_5 w^2/n} = c_4 \sum_{w=1}^{\infty} e^{-c_5 w^2/n}.$$

The sum $\sum_{w=1}^{\infty} e^{-c_5 w^2/n}$ resembles a Riemann sum and can be approximated by an integral for large n. We have

$$\sum_{w=1}^{\infty} e^{-c_5 w^2/n} \leq \int_0^{\infty} e^{-c_5 x^2/n}\, dx.$$

To compute this integral, we use the substitution $u = \sqrt{\frac{c_5}{n}} x$, which implies $x = \sqrt{\frac{n}{c_5}} u$ and $dx = \sqrt{\frac{n}{c_5}}\, du$. Substituting into the integral gives

$$\int_0^{\infty} e^{-c_5 x^2/n}\, dx = \sqrt{\frac{n}{c_5}} \int_0^{\infty} e^{-u^2}\, du = \sqrt{\frac{n}{c_5}} \cdot \frac{\sqrt{\pi}}{2}.$$

Therefore,

$$\int_0^{\infty} e^{-c_5 x^2/n}\, dx = \sqrt{\frac{n}{c_5}} \cdot \frac{\sqrt{\pi}}{2}.$$

Thus, the sum can be approximated as

$$\sum_{w=1}^{\infty} e^{-c_5 w^2/n} \sim \sqrt{\frac{n}{c_5}} \cdot \frac{\sqrt{\pi}}{2}.$$

Substituting this back into the bound for $\mathbb{E}[W(T_n)]$, we find

$$\mathbb{E}[W(T_n)] \leq c_4 \cdot \sqrt{\frac{n}{c_5}} \cdot \frac{\sqrt{\pi}}{2}.$$

This shows that

$$\mathbb{E}[W(T_n)] = \mathcal{O}(\sqrt{n}).$$ □

3.3. The height

Definition 3.9. The height of a tree T, denoted by $H(T)$, is the maximum depth of any node in the tree.

$$\mathrm{H}(T) = \max_{v \in V(T)} d(v).$$

A lexicographic depth-first search (DFS) is a linear time algorithm for ordering the vertices of a labelled graph. At each node, its children are visited in lexicographical order. The children of the first children are explored before going to its siblings. This idea is applied recursively for each vertex. Let Q_i^d be the number of nodes in the stack of the DFS at the i-th step. We define $Q_0^d = 1$. At each step, we get rid of a vertex from the queue and add its children, which we read in the lexicographical order. Hence, since all individuals reproduce with the same probability distribution, the following recursion is clear:

$$Q_i^d = Q_{i-1}^d - 1 + \chi_i.$$

The reverse-lexicographic depth-first search is a variation of the standard depth-first search (DFS) where, instead of visiting the children of a node in lexicographical order (smallest first), we visit the children in reverse lexicographical order (largest first). We will denote by Q_i^r the number of nodes in the queue of the algorithm at the i-th step.

One can easily observe that in a uniformly random tree T_n, the labels of the vertices do not affect the probability distribution of the labeled tree. While BFS explores nodes level by level and DFS goes deeper first, lexicographic and reverse-lexicographic DFS are merely different ways of ordering the visits to the nodes. Therefore, the resulting labelings observed along these procedures follow the same distribution.

Now, we will stand out several preliminary lemmas:

Lemma 3.10. *Let P be the unique path in a tree from a vertex which has height h to the root. The expected value of nodes in P that have more than one child is $h \cdot q_1$, where $q_1 = 1 - \mathbb{P}(\chi = 1)$.*

By the previous lemma, the expected number of nodes in P that have exactly one child is $h \cdot p_1$, where $p_1 = \mathbb{P}(\chi = 1)$.

3.4. A modified Bienaymé–Galton–Watson tree

To complete the study of the height of a random tree, we will need to introduce a new concept.

Let $\hat{\chi}$ be a random variable with the size-biased distribution

$$\mathbb{P}(\hat{\chi} = m) = m \cdot \mathbb{P}(\chi = m).$$

It is clear that this is a probability distribution since $\sum_m \mathbb{P}(\tilde{\chi} = m) = \mathbb{E}[\chi] = 1$, and that $\hat{\chi} \geq 1$.

Let, for $k \geq 1$, $T^{(k)}$ be the modified BGW defined as follows. It has two different types of nodes: normal and mutant. Normal constitute the offspring of χ, while mutant nodes have offspring according to independent copies of $\hat{\chi}$. All children of normal nodes are also normal. Exactly one child of each mutant node is chosen at random and it is called its heir. If this heir has depth less than k, it is mutant. If not, it is normal. Hence, there are exactly k mutant nodes, which together with the heir v^* of the last mutant node, form a path from the root to v^* at depth k. This path is what we call the spine γ of $\hat{T}^{(k)}$.

An equivalent construction of the modified BGW tree can be described as follows:

1. **Construct the spine:**
    - The spine is a path consisting of k nodes, starting from the root at depth 0 and ending at a node at depth k.
    - These nodes correspond to the mutant nodes in the original construction.
2. **Attach independent subtrees to the nodes of the spine:**
    - For each node along the spine, except the last one at depth k, attach a number of independent subtrees. The number of such subtrees is distributed as $\hat{\chi} - 1$, meaning it follows the distribution $\hat{\chi}$ reduced by 1.
    - For the node of the spine at depth k, attach independent subtrees according to the distribution χ, which corresponds to the behavior of normal nodes.

Hence, the spine forms the central path from the root to depth k, while the subtrees, attached independently, reflect the probabilistic structure of the original model.

Now, we aim to study the probability of obtaining a particular tree T from a modified BGW tree $\hat{\mathcal{T}}^{(k)}$. To start with it, it is not difficult to find the probability that a given mutant node has m children and one of them is selected as heir. This is:

$$\frac{1}{m}\mathbb{P}(\hat{\chi} = m) = \mathbb{P}(\chi = m).$$

Hence, since every normal node has distribution χ and the process consisting of taking a mutant node with m children and selecting one of them to be mutant also distributes as χ, the following is clear:

$$\mathbb{P}(\hat{\mathcal{T}}^{(k)} = T \text{ with } \gamma \text{ as spine}) = \prod_v \mathbb{P}(\chi = d_v) = \mathbb{P}(X = T).$$

Then, the probability of getting a particular tree T from a modified BGW with a fixed spine is the same of obtaining this T from the BGW tree. Since for every node v at depth k, there exists one unique path from v to the root, there is a bijection between nodes at depth k and spines. Therefore, summing for all nodes at depth k:

$$\mathbb{P}(\hat{\mathcal{T}}^{(k)} = T) = Z_k(T) \cdot \mathbb{P}(\mathcal{T} = T).$$

In conclusion, $\hat{\mathcal{T}}^{(k)}$ has the distribution of $\mathcal{T}$ biased by Z_k, which is the size of the k-th generation.

Returning to the broader discussion, with these preliminaries in place, we now have the necessary tools to prove the following.

Theorem 3.11. *Let χ be a discrete random variable taking values in nonnegative integers. Suppose that $\mathbb{E}[\chi] = 1$ and $0 < \mathrm{Var}(\chi) < \infty$. Let T_n be a BGW(χ) tree conditional on having n nodes. Then, for all $n \geq 1$ and $h \geq \sqrt{n}$,*

$$\mathbb{P}(H(T_n) \geq h) \leq C_2 e^{-c_2 h^2/n},$$

where $C_2 > 0$ and $c_2 > 0$.

Proof. We will follow the proof approach provided in [1]. Let h be the height of T_n. We may assume that $h \in \mathbb{Z}$. We will base the proof on the next observation. Since we have proved that the width is expected to be $\sqrt{n}$, we can suppose there is a vertex $v \in V(T_n)$ with "large" height. Hence, there are two possible cases of obtaining a tree with height h: either there are many edges leaving the unique path P from v to the root, where v is a vertex from the h-th level, or there are many of the ancestors of v with just one child.

In the first case, the objective is to bound $\max(Q_j^d, Q_k^r)$ and connect this bound with the probabilistic structure of the tree. The quantities Q_j^d and Q_k^r measure how many vertices are simultaneously active during the lexicographic and reverse-lexicographic depth-first explorations, respectively. Whenever a node on the path from a vertex at height h to the root has more than one child, all of its extra children are added to the exploration data structure and remain there until they are processed. This causes Q_j^d or Q_k^r to increase, and their maximum size thus reflects the cumulative effect of branching along the path. Therefore, the height h of the tree can be directly related to $\max(Q_j^d, Q_k^r)$: if the tree has height h, then there must exist an index j with $Q_j^d = h$ or an index k with $Q_k^r = h$.

Let $p_1 = \mathbb{P}(\chi = 1)$ and let $q_1 = 1 - p_1$. Let $v \in V(T_n)$ such that $h(v) = h$. Let j (resp. k) be the index of v in lexicographic (resp. reverse-lexicographic) order. Let X be the number of nodes which have more than one child in P. Each ancestor of v with more than one child contributes to at least one unit to Q_j^d or to Q_k^r. We distinguish two cases, either $\max(Q_j^d, Q_k^r) \geq \frac{q_1}{3}h$ or $\max(Q_j^d, Q_k^r) < \frac{q_1}{3}h$. In the second case, by the above remark, the number of ancestors of v with exactly one child is at least $(1-2q_1/3)h = (p_1+q_1/3)h$.

Now, it is not useful to think about the queues because when the algorithm processes a node which has just one child, the size of the queue does not increase, so it is not a good representation for the height of the tree. Let $S(h)$ be the set of trees T with $|T| = n$ and containing a node v such that $h(v) = h$ that has at least $(p_1 + q_1/3)h$ ancestors in P with exactly one child. Then, let $\alpha := \{T_n \in S\} = \bigcup_{T\in S}\{T_n = T\}$.

Then, we can apply these two cases described to the calculus of the following probability. The first two terms correspond to the first case, while the remaining term corresponds to the second case:

$$\begin{aligned}\mathbb{P}(H(T_n) \geq h) &\leq \mathbb{P}\left(\max_j Q_j^d \geq \frac{q_1}{3}h\right) + \mathbb{P}\left(\max_k Q_k^r \geq \frac{q_1}{3}h\right) + \mathbb{P}(\alpha)\\ &= 2\mathbb{P}\left(\max_i Q_i \geq \frac{q_1}{3}h\right) + \mathbb{P}(\alpha) \leq \ C_{11}e^{-c_{11}h^2/n} + \mathbb{P}(\alpha),\end{aligned}$$

where the last inequality has been seen in the proof of Theorem 3.7.

Then, we only need to bound $\mathbb{P}(\alpha)$:

$$\begin{aligned}\mathbb{P}(\mathcal{T} \in S) &= \sum_{T\in S}\mathbb{P}(\mathcal{T} = T) = \sum_{T\in S}\mathbb{P}(\hat{T}^{(h)} = T \text{ with } \gamma_T \text{ as spine})\\ &= \mathbb{P}\left(\bigcup_{T\in S}\{\hat{T}^{(h)} = T \text{ with } \gamma_T \text{ as spine}\}\right) \leq \mathbb{P}\left(\sum_{i=0}^{h-1}\mathbf{1}_{\hat{\chi}_i=1} \geq (p_1 + q_1/3)h\right).\end{aligned}$$

To justify the last inequality, recall the construction of $\hat{T}^{(h)}$: the spine γ consists of the h mutant nodes from the root to depth h, and for each $i = 0, \dots, h-1$ the number of offspring of the i-th mutant is distributed as $\hat{\chi}_i$ (the size-biased distribution), one of whose children is then chosen uniformly to continue the spine.

There is a one-to-one correspondence between the property "the i-th vertex on the spine has exactly one child in the realised tree T" and the event "$\hat{\chi}_i = 1$ in the construction of $\hat{T}^{(h)}$": if the i-th mutant has exactly one child, then necessarily that child is the heir (so $\hat{\chi}_i = 1$), and conversely, if $\hat{\chi}_i = 1$, then the i-th spine vertex has exactly one child in the realised tree.

Consequently, whenever $\hat{T}^{(h)}$ equals some $T \in S$ with spine γ_T, the number of spine vertices having exactly one child is at least $(p_1 + q_1/3)h$ by the definition of S. Therefore the event

$$\bigcup_{T\in\mathcal{S}}\{\hat{T}^{(h)} = T \text{ with } \gamma_T \text{ as spine}\}$$

is contained in the event

$$\left\{\sum_{i=0}^{h-1}\mathbf{1}_{\{\hat{\chi}_i=1\}} \geq (p_1 + q_1/3)h\right\},$$

which yields the displayed inequality.

Since $\mathbf{1}_{\hat{\chi}_i=1}$ are Bernoulli(p_1), by Lemma 3.6, and taking $t = q_1 h/3$, $V = p_1 q_1 h$, $b = q_1$, we have the following bound:

$$\mathbb{P}\left(\sum_{i=0}^{h-1} \mathbf{1}_{\hat{\chi}_i=1} \geq (p_1 + q_1/3)h\right) \leq \exp\left(-\frac{(q_1 h/3)^2}{2p_1 q_1 h + 2q_1^2 h/9}\right) = \exp\left(-\frac{h}{18p_1/q_1 + 2}\right).$$

Furthermore, the following equality is clear:

$$\mathbb{P}(\mathcal{T} \in S) = \mathbb{P}(\alpha) \cdot \mathbb{P}(|\mathcal{T}| = n).$$

By Lemma 3.5, $\mathbb{P}(|\mathcal{T}| = n) \geq n^{-3/2}$ and we can establish an upper bound for $\mathbb{P}(\delta)$.

$$\mathbb{P}(\delta) = \frac{P(\mathcal{T} \in S)}{P(|\mathcal{T}| = n)} \leq C_{12} n^{3/2} \exp\left(-\frac{h}{18p_1/q_1 + 2}\right) \leq C_{12} n^{3/2} \exp\left(\frac{h}{18p_1/q_1 + 2}\frac{h}{\sqrt{n}}\right) \leq C_{13} e^{-c_{12} h^2/n},$$

for all $h \geq \sqrt{n}$. Taking everything into account, what we have is the following bound for the height of a BGW tree:

$$\mathbb{P}(H(\mathcal{T}_n) \geq h) \leq C_{11} e^{-c_{11} h^2/n} + C_{13} e^{-c_{12} h^2/n}.$$

Now, taking

$$C_2 := \max\{C_{11}, C_{13}\}, \quad c_2 := \min\{c_{11}, c_{12}\} \Longrightarrow \mathbb{P}(H(\mathcal{T}_n) \geq h) \leq C_2 e^{-c_2 h^2/n}. \qquad \square$$

Having proved that, we left the following result without proof due to its analogy with Theorem 3.8.

Theorem 3.12. *Let χ be a discrete random variable taking values in nonnegative integers. Let T_n be a* BGW(χ) *tree conditional on having n nodes. Suppose that $\mathbb{E}[\chi] = 1$ and $0 < \mathrm{Var}(\chi) < \infty$. Then, for all $n \geq 1$,*

$$\mathbb{E}[H(T_n)] = \mathcal{O}(\sqrt{n}).$$

References

[1] L. Addario-Berry, L. Devroye, S. Janson, Sub-Gaussian tail bounds for the width and height of conditioned Galton–Watson trees, *Ann. Probab.* **41(2)** (2013), 1072–1087.

[2] D. Aldous, The continuum random tree. I, *Ann. Probab.* **19(1)** (1991), 1–28.

[3] V.F. Kolchin, *Random Mappings*, Translated from the Russian, With a foreword by S. R. S. Varadhan, Transl. Ser. Math. Engrg., Optimization Software, Inc., Publications Division, New York, 1986.

REPORTS@SCM
AN ELECTRONIC JOURNAL OF THE
SOCIETAT CATALANA DE MATEMÀTIQUES

On the basins of attraction of root-finding algorithms

David Rosado Rodríguez
Universitat de Barcelona
rosadodav4@gmail.com

Resum *(CAT)*
Els algoritmes de cerca d'arrels s'han utilitzat per resoldre numèricament equacions no lineals de la forma $f(x) = 0$. Aquest article estudia la dinàmica de la família parametritzada de Traub $T_{p,\delta}$ aplicada a polinomis, que abasta des del mètode de Newton ($\delta = 0$) fins al de Traub ($\delta = 1$). Ens centrem en propietats topològiques de les conques immediates d'atracció dels punts fixos finits, especialment la seva connectivitat i el fet de ser acotada o no. Aquests fets són clau per identificar condicions inicials universals que assegurin la convergència a totes les arrels de p.

Abstract *(ENG)*
Root-finding algorithms have historically been used to numerically solve nonlinear equations of the form $f(x)=0$. This paper studies the dynamics of the parameterized Traub family $T_{p,\delta}$ applied to polynomials, ranging from Newton's method ($\delta = 0$) to Traub's method ($\delta = 1$). We focus on topological properties of the immediate basins of attraction of the finite fixed points, especially simple connectivity and unboundedness, which are key to identifying a universal set of initial conditions ensuring convergence to all roots of p.

Keywords: *dynamical systems, root-finding algorithms, Newton's method, Traub's method.*
MSC (2020): *Primary 30D05, 37F10, 37F46.*

Received: *April 27, 2025.*
Accepted: *August 8, 2025.*

1. Introduction

Solving non-linear equations of the form $f(x) = 0$ is a common challenge in various scientific fields, spanning from biology to engineering. When algebraic manipulation is not feasible, iterative methods become necessary to determine solutions. Among these, Newton's method stands out as a widely used approach, relying on the linearization of $f(x)$. Its iterative scheme is given by:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \quad n \geq 0.$$

Nevertheless, numerous numerical methods have proven to be efficient when they converge, including the one considered here, Traub's method. While Newton's method exhibits quadratic convergence for simple roots of a polynomial when the initial guess is sufficiently close to the root, Traub's method achieves cubic (local) convergence. This method belongs to a parametric family of iterative schemes, first introduced in [6, 12], known as the *damped Traub's family*. Its iterative formula is given by:

$$x_{n+1} = y_n - \delta \frac{f(y_n)}{f'(x_n)}, \quad n \geq 0,$$

where $y_n = x_n - \frac{f(x_n)}{f'(x_n)}$ represents a Newton step, and δ is the damping parameter. Notably, setting $\delta = 0$ recovers Newton's method, while $\delta = 1$ corresponds to Traub's method. It is important to mention that each iteration of Traub's method involves additional computations compared to Newton's method. Although the question can be explored in other settings, here we will focus on the case where $p(z) = 0$, $z \in \mathbb{C}$.

Roughly speaking, when we have a good estimate of the solutions to the equation $p(z) = 0$, iterative methods tend to work well. However, challenges arise when the number of solutions of f is large or when we lack control over these solutions. This is particularly problematic when selecting initial conditions to initiate the algorithm. In such situations, the study of dynamical systems becomes valuable. By examining the topological characteristics of the immediate basins of attraction associated with the solutions of $p(z) = 0$, we can gain valuable insights and aid in addressing these challenges. An illustration of this is provided in [8]. In their work, the authors used some topological results of the basins of attraction to construct a universal and explicit set of initial conditions denoted as $\mathcal{S}_d$. This set, depending only on the polynomial's degree, allows Newton's method to find all roots of a polynomial. The existence of the set $\mathcal{S}_d$ is guaranteed by the following key properties of the immediate basins of attractions for the Newton's method (first proved by Przytycki [10] and later generalized by Shishikura [11]).

Theorem 1.1. *Let p be a polynomial of degree $d \geq 2$. Assume that $p(\alpha) = 0$ and let N_p be the corresponding Newton's map. Then, the immediate basin of attraction of α, denoted as $\mathcal{A}^*(\alpha)$, is a simply connected, unbounded set.*

A natural question that comes up now is whether we can create a set similar to $\mathcal{S}_d$ for Traub's method. If this were possible, it would provide a way to find all the roots of a polynomial with improved convergence speed. Specifically, as previously noted, for simple roots of the polynomial, the local convergence order would be cubic instead of quadratic, leading to faster convergence. To achieve this, proving an equivalent to Theorem 1.1 for Traub's method, will provide the necessary tools for building the $\mathcal{S}_d$ like-set. In a recent study [3], Theorem 1.1 was proved for Traub's method under certain additional assumptions. To be precise, the researchers successfully established the following theorem:

Theorem 1.2. *Let p be a polynomial of degree $d \geq 2$. Assume that p satisfies one of the following conditions:*

(i) $d = 2$, *or*

(ii) *it can be written in the form* $p_{n,\beta}(z) := z^n - \beta$ *for some* $n \geq 3$ *and* $\beta \in \mathbb{C}$.

Suppose that $p(\alpha) = 0$ *and consider damped Traub's map* $T_{p,\delta}(z) := N_p(z) - \delta \frac{p(N_p(z))}{p'(z)}$ *with* $\delta \in [0, 1]$. *Then* $\mathcal{A}^*_\delta(\alpha)$ *is a simply connected unbounded set.*

This article explores the damped Traub's method as a root-finding algorithm. We provide background to understand the proof of Theorem 1.2 and present two results that bring us closer to our goal: proving an equivalent result to Theorem 1.1 for Traub's method. Specifically, we establish the following result:

Theorem A. *Let p be a polynomial of degree $d \geq 2$. Assume that $p(\alpha) = 0$ and let $T_{p,\delta}$ be the corresponding damped Traub's map. Then, for δ close enough to zero,* $\mathcal{A}^*_{T_{p,\delta}}(\alpha)$ *is an unbounded set.*

Moreover, we analyze the behavior of Traub's method for the polynomial family $p_d(z) = z(z^d - 1)$. Notably, for Halley's root-finding algorithm, the basin of attraction of $z = 0$ is bounded when $d = 5$, but we establish the following result:

Theorem B. *Let* $p_d(z) = z(z^d - 1)$. *Then,* $\mathcal{A}^*_{T_{p_d,1}}(0)$ *is an unbounded set for every integer* $d > 0$.

2. An introduction to holomorphic dynamics

Let us denote $\hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ the *extended complex plane* or *Riemann Sphere*.

Definition 2.1. Let $R \colon \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ be a rational map. A point $z = z_0$ is a *fixed point* if $R(z_0) = z_0$ (resp. *periodic of period p* if $R^p(z_0) = z_0$ for some $p \geq 1$ and $R^n(z_0) \neq z_0$ for all $n < p$). The *multiplier* of z_0 is $\lambda = R'(z_0)$ (resp. $\lambda = (R^p)'(z_0)$).

The character of the fixed or periodic points can be determined by using the multiplier. In fact, the fixed or periodic point z_0 is *attracting* if $|\lambda| < 1$ (*superattracting* if $\lambda = 0$), *repelling* if $|\lambda| > 1$ and *indifferent* if $|\lambda| = 1$.

Definition 2.2. Let $R \colon \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ be a rational map and $z_0 \in \hat{\mathbb{C}}$ be an attracting fixed point of R. We define the *basin of attraction* of z_0 as

$$\mathcal{A}_R(z_0) := \mathcal{A}(z_0) = \{z \in \hat{\mathbb{C}} \colon R^n(z) \xrightarrow{n\to\infty} z_0\}.$$

We denote by $\mathcal{A}^*(z_0)$ the connected component of $\mathcal{A}(z_0)$ containing z_0, and we refer to it as the *immediate basin of attraction*.

In what follows we omit the dependence with respect to the rational map under consideration, unless it is mandatory. It is easy to see that $\mathcal{A}(z_0)$ is an open set containing z_0. There is a vast body of results on this topic, and for a general overview, many excellent references are available; see, for instance, [2, 9, 5]. To conclude this chapter, we present a theorem that will be useful later. This theorem state that, in a neighbourhood of an attracting fixed point, the map *looks like* $g(\zeta) = \lambda\zeta$.

Theorem 2.3 (**Koenigs linearization Theorem**). *Let $z_0 \in \mathbb{C}$, U neighbourhood of z_0 and $f : U \to \mathbb{C}$ be a holomorphic function such that z_0 is an attracting fixed point with multiplier $0 < |\lambda| < 1$. Then there is a conformal map $\zeta = \phi(z)$ of a neighbourhood of z_0 onto a neighbourhood of z_0 which conjugates f to the linear function $g(\zeta) = \lambda\zeta$. The conjugating function is unique, up to multiplication by a nonzero scale factor.*

3. Local dynamics of the family $T_{p,\delta}$

Recall that if p is a polynomial of degree $d \geq 2$, the damped Traub's map applied to p is defined as

$$T_{p,\delta} = N_p(z) - \delta \frac{p(N_p(z))}{p'(z)},$$

where N_p is the Newton's map and $\delta \in \mathbb{C}$. For our purposes, it will suffice to consider $\delta \in [0, 1]$. Notice that, setting $\delta = 0$ recovers the well-known Newton's method. The map N_p is the *universal* root-finding algorithm and it satisfies the following key global dynamical properties:

Proposition 3.1. *Let p be a polynomial of degree $d \geq 2$. The following properties regarding the Newton's map hold:*

(i) A point $z = \alpha$ is a root of p if and only if it is a fixed point of N_p.

(ii) The simple roots of p are superattracting fixed points of N_p, while multiple roots are attracting fixed points of N_p.

(iii) The point $z = \infty$ is the only repelling fixed point of N_p.

Proof. (i) and (ii) are straightforward computations. For (iii), observe that $N_p(\infty) = \lim_{z\to\infty} N_p(z) = \infty$, so $z = \infty$ is a fixed point. To see its nature, consider the transformation $\phi\colon U \to V$ such that $\phi(z) = 1/z$, where U is a neighbourhood of $z = \infty$ and V is a neighbourhood of $z = 0$. The conjugate map is then $\widetilde{N_p}(z) = \phi(N_p(\phi^{-1}(z))) = \frac{1}{N_p(1/z)}$, so studying the behaviour of $\widetilde{N_p}$ at $z = 0$ reveals the character of $z = \infty$ in the original system. □

Leveraging the properties of Newton's method, particularly noting that $z = \infty$ is the only repelling fixed point, Theorem 1.1 can be established. Details of the proof can be found in [1]. For $\delta \neq 0$, some of the properties that hold for N_p remain the same, while others change slightly. Let us summarize them:

Proposition 3.2. *Let p be a polynomial of degree $d \geq 2$ and $\delta \in (0, 1]$. The following properties regarding the damped Traub's map hold:*

(i) If $z = \alpha$ is a root of p, then $z = \alpha$ is a fixed point of $T_{p,\delta}$. The converse is not necessarily true.

(ii) The simple roots of p are superattracting fixed points of $T_{p,\delta}$, while multiple roots are attracting fixed points of $T_{p,\delta}$.

(iii) The point $z = \infty$ is a repelling fixed point of $T_{p,\delta}$.

The complete and detailed proof can be found in [3]. Notice that, finite fixed points of the method do not necessarily correspond to zeros of the polynomial, as is the case with Newton's method; see Figure 1 for a visual illustration. With all this information, a recent study [3] successfully established Theorem 1.2.

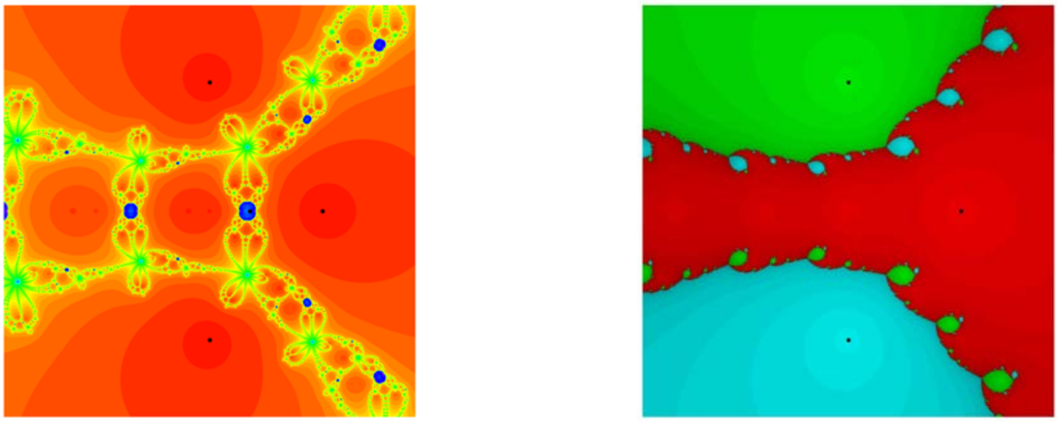

Figure 1: On the left, we illustrate the dynamical plane of Traub's method applied to the cubic polynomial $P(z) = (z^2 + 0.25)(z - 0.439)$. Basins of attraction corresponding to roots of the polynomial are shown in orange. It is notable that $T_{p,1}$ exhibits an attracting fixed point located at $\zeta \approx 0.155$, whose basin is depicted in blue, that does not correspond to any root of P. On the right, we present the dynamical plane of Newton's method applied to P. Here, it is evident that there are no fixed points other than the roots.

4. The method as a singular perturbation

In this chapter, we prove the unboundedness of immediate basins of attraction when $\delta \approx 0$. For small δ, the damped Traub's method acts as a singular perturbation of Newton's method. A *singular perturbation* refers to a base family (with well-understood dynamics) combined with a local perturbation that increases the map's degree and enriches its dynamics. This perturbation affects only certain regions of the dynamical plane when the parameter is small [7]. Here, Newton's method is the base family, and $p(N_p(z))/p'(z)$ is the perturbation. Notice that the singular perturbation occurs over the Julia set, as it involves adding additional preimages of $z = \infty$ to the zeros of $p'(z)$. To establish the main result of the section, we will first present some auxiliary results.

Lemma 4.1. *Let $p(z) = a_d z^d + \cdots + a_1 z + a_0$ be a polynomial of degree $d \geq 2$. Let q_j be the zeros of $p'(z) = 0$, i.e., the poles of both the damped Traub's map, $T_{p,\delta}$, and Newton's map, N_p. Consider the compact $K = \overline{D(0,R)} \setminus \cup_j D(q_j, \varepsilon)$ where $R > 0$ and $\varepsilon > 0$ are positive fixed constants. Then, for every $z \in K$, there exists a constant $C_{R,\varepsilon}$ such that $|p(N_p(z))/p'(z)| \leq C_{R,\varepsilon}$.*

Proof. Let $z \in K$. There exists a positive value $\eta_\epsilon > 0$ such that $|p'(z)| > \eta_\epsilon$. Moreover, since $|z| < R$, $|p(z)| \leq |a_d|R^d + \cdots + |a_1|R + |a_0| := R'$. Hence,

$$|N_p(z)| \leq |z| + \left|\frac{p(z)}{p'(z)}\right| \leq R + \frac{R'}{\eta_\epsilon} := M.$$

Therefore,

$$\left|\frac{p(N_p(z))}{p'(z)}\right| \leq \frac{|a_d N_p(z)^d| + \cdots + |a_1 N_p(z)| + |a_0|}{\eta_\epsilon} \leq \frac{|a_d|M^d + \cdots + |a_1|M + |a_0|}{\eta_\epsilon} := C_{R,\varepsilon}. \qquad \square$$

Lemma 4.2. *Let p be a polynomial of degree $d \geq 2$. Let q_j be the zeros of $p'(z) = 0$, i.e., the poles of both the damped Traub's map, $T_{p,\delta}$, and Newton's map, N_p, and let $z = \alpha$ be a zero of p, i.e., an attracting fixed point for both N_p and $T_{p,\delta}$. Consider the compact $K = \overline{D(0,R)} \setminus \cup_j D(q_j, \varepsilon')$ where $R > 0$ and $\varepsilon' > 0$ are positive fixed constants such that $\alpha \in K$. Then, the following statements hold:*

(i) There exists a compact $K' \subset K$ such that $K' \subset \mathcal{A}^*_{N_p}(\alpha)$, $\alpha \in K'$ and $\partial K' \cap \partial K \neq \emptyset$, satisfying that for every $z \in K'$, there is a unique $M \in \mathbb{N}$ such that: $\forall \varepsilon > 0$, $N_p^M(z) \in D(\alpha, \varepsilon/2)$.

(ii) For the given $\varepsilon > 0$ and for δ small enough, the following property holds: $\forall z \in K'$, $|N_p^M(z) - T_{p,\delta}^M(z)| < \varepsilon/2$. In particular, $T_{p,\delta}^M(z) \in D(\alpha, \varepsilon)$.

Proof. (i) The existence of such a compact is guaranteed by the fact that $\mathcal{A}^*_{N_p}(\alpha)$ is an open set, unbounded and simply connected. Since $z = \alpha$ is an attracting fixed point for N_p, the existence of $M \in \mathbb{N}$ is also guaranteed.

(ii) To prove the result, let us first establish the following claim: For δ small enough,

$$\forall r > 0, \ \exists \rho > 0 \text{ such that if } |z_1 - z_2| < \rho \implies |N_p(z_1) - T_{p,\delta}(z_2)| < r.$$

To prove the claim, observe that, using Lemma 4.1 in the last inequality,

$$|N_p(z_1) - T_{p,\delta}(z_2)| \leq |N_p(z_1) - N_p(z_2)| + \delta \left| \frac{p(N_p(z))}{p'(z)} \right| \leq |N_p(z_1) - N_p(z_2)| + \delta C_{R,\varepsilon'}.$$

Hence, since Newton's map is continuous in K (in particular it is also in K'), there exists $\rho > 0$ such that if $|z_1 - z_2| < \rho$, then $|N_p(z_1) - N_p(z_2)| < r/2$. Setting $\delta = \frac{r}{2C_{R,\varepsilon'}}$, we obtain the desired bound.

Now, let $z \in K$. To prove the result, we proceed as follows:

1. Using the claim with $z_1 := N_p^{M-1}(z)$ and $z_2 := T_{p,\delta}^{M-1}(z)$, there exists $\eta_M > 0$ and $\delta_M > 0$ such that if $|N_p^{M-1}(z) - T_{p,\delta}^{M-1}(z)| < \eta_M$, then $|N_p^M(z) - T_{p,\delta_M}^M(z)| < \varepsilon/2$.
2. Iterating the algorithm, we obtain sequences $\{\eta_{M-i}\}_{i=0}^{M-3}$, $\{\delta_{M-i}\}_{i=0}^{M-3}$ satisfying that if $|N_p^{M-i-1}(z) - T_{p,\delta}^{M-i-1}(z)| < \eta_{M-i}$, then $|N_p^{M-i}(z) - T_{p,\delta_i}^{M-i}(z)| < \eta_{M-i+1}$.
3. We conclude the algorithm with the existence of $\eta_2 > 0$ and $\delta_2 > 0$ such that if $|N_p(z) - T_{p,\delta_2}(z)| < \eta_2$, then $|N_p^2(z) - T_{p,\delta_2}^2(z)| < \eta_3$.

Finally, to ensure that $|N_p(z) - T_{p,\delta}(z)| < \eta_2$, we just need to take $\delta_1 = \frac{\eta_2}{C_{R,\varepsilon'}}$. Therefore, taking $\delta = \min\{\delta_1, \dots, \delta_M\}$, we obtain that for every $z \in K$: $|T_{p,\delta}^M(z) - N_p^M(z)| < \varepsilon/2$. □

Theorem A. *Let p be a polynomial of degree $d \geq 2$. Assume that $p(\alpha) = 0$ and let $T_{p,\delta}$ be the corresponding damped Traub's map. Then, for δ close enough to zero, $\mathcal{A}^*_{T_{p,\delta}}(\alpha)$ is an unbounded set.*

Proof. First, observe that for δ close enough to zero (indeed for every $\delta \in [0,1]$), $z = \infty$ is a repelling fixed point for $T_{p,\delta}$. By Koenigs linearization Theorem, in a neighborhood of $z = \infty$, say $D(\infty, \varepsilon)$, $T_{p,\delta}$ is locally conjugated to $g(\zeta) = \lambda\zeta$, where λ is the multiplier of $z = \infty$. Notice that, if $\lambda \in \mathbb{C}$, since $|\lambda| > 1$, points near $z = \infty$ tend to move away in a spiral shape, and if $\lambda \in \mathbb{R}$, since $|\lambda| > 1$, points near $z = \infty$ tend to move outward in a radial manner.

Let us define $R := \frac{1}{\varepsilon}$ and consider the compact $K := \overline{D(0,R)} \setminus \cup_j D(q_j, \varepsilon')$ where q_j are the poles of $T_{p,\delta}$, i.e., the zeros of $p'(z) = 0$, and $\varepsilon' > 0$ is a positive fixed constant. We can assume that $\alpha \in K$. If not, we can choose a smaller value for ε (increasing the value of R) to ensure that $\alpha \in K$, making the neighborhood where the Koenigs coordinates apply smaller. Since $z = \alpha$ is an attracting fixed point for both N_p and $T_{p,\delta}$, there exists $\eta_1, \eta_2 > 0$ such that $D(\alpha, \eta_1) \subset \mathcal{A}^*_{N_p}(\alpha)$ and $D(\alpha, \eta_2) \subset \mathcal{A}^*_{T_{p,\delta}}(\alpha)$. Setting $\eta = \min\{\eta_1, \eta_2\}$, we have that $D(\alpha, \eta) \subset \mathcal{A}^*_{N_p}(\alpha) \cap \mathcal{A}^*_{T_{p,\delta}}(\alpha)$. According to Lemma 4.2(i), there exists a compact $K' \subset K$ such that $K' \subset \mathcal{A}^*_{N_p}(\alpha)$, $\alpha \in K'$ and $\partial K' \cap \partial K \neq \emptyset$, satisfying that for every $z \in K'$, there is a unique $M \in \mathbb{N}$ such that, for every $z \in K'$, $N_p^M(z) \in D(\alpha, \eta/2) \subset D(\alpha, \eta)$. Moreover, since the basins of attraction of Newton's method are unbounded and simply connected, there exists a ray τ connecting the fixed point $z = \alpha$ and $z = \infty$, included in $\mathcal{A}^*_{N_p}(\alpha)$. This ray can be chosen such that its restriction to K is included in K'. From now own, any reference to τ will indicate the ray extending from the point $z = \alpha$ to the boundary of the set K. Then, according to Lemma 4.2(ii), for δ small enough and $z \in K'$, $T^M_{p,\delta}(z) \in D(\alpha, \eta)$, indicating that $z \in \mathcal{A}^*_{T_{p,\delta}}(\alpha)$. Then, either $\tau \subset K' \subset \mathcal{A}^*_{T_{p,\delta}}(\alpha)$ or there exists $w \in \mathcal{J}(T_{p,\delta}) \cap K'$. In the last case, since $w \in K'$, $T^M_{p,\delta}(w) \in D(\alpha, \eta)$, in contradiction with $w \in \mathcal{J}(T_{p,\delta})$. Therefore, $\tau \subset K' \subset \mathcal{A}^*_{T_{p,\delta}}(\alpha)$.

By construction, observe that $\partial D(0,R) = \partial D(\infty, \varepsilon)$, hence, the ray τ, which ends at $\partial D(0,R)$, connects with the spiral (or the line in case $\lambda \in \mathbb{R}$) that extends towards $z = \infty$. Thus, we found a ray that connects the fixed point $z = \alpha$ to $z = \infty$, which is contained within $\mathcal{A}^*_{T_{p,\delta}}(\alpha)$. This proves that the immediate basin of attraction for the damped Traub's method is unbounded when $\delta \approx 0$. $\square$

5. Traub's method applied to $z(z^d - 1)$

Now, we aim to examine Traub's method applied to the family $p_d(z) = z(z^d - 1)$. This family is particularly interesting because, for Halley's root-finding algorithm, it was found that for $d = 5$, the immediate basin of attraction of $z = 0$ is bounded. Therefore, proving that this is not the case for Traub's method would support the conjecture that the immediate basins of attraction of Traub's method are unbounded. We have been able to prove that the immediate basin of attraction of $z = 0$ is unbounded for every d. To establish the result, we will first present an auxiliary result.

Lemma 5.1. *The semi-lines $z = re^{\frac{(2k+1)\pi i}{d}}$, $r > 0$ and $k = 0, 1, \ldots, d-1$, are forward invariant under $T_{p_d,1}$.*

First of all, observe that

$$T_{p_d,1}(z) = N_{p_d}(z) - \frac{p_d(N_{p_d}(z))}{p_d'(z)} = \frac{d(d+1)z^{2d+1}[(d+1)z^d - 1]^d - [dz^{d+1}]^{d+1}}{[(d+1)z^d - 1]^{d+2}}.$$

Hence, since $e^{(2k+1)\pi i} = -1$, a straightforward computation reveals that $T_{p_d,1}(re^{\frac{(2k+1)\pi i}{d}}) = e^{\frac{(2k+1)\pi i}{d}} R_d(r)$, where

$$R_d(r) := \frac{d(d+1)r^{2d+1}[(d+1)r^d + 1]^d - d^{d+1}r^{(d+1)^2}}{[(d+1)r^d + 1]^{d+2}}.$$

Theorem B. *Let $p_d(z) = z(z^d - 1)$. Then, $\mathcal{A}^*_{T_{p_d,1}}(0)$ is an unbounded set for every integer $d > 0$.*

Proof. Consider only the semi-lines that do not cross the dth roots of unity, i.e., $z = re^{\frac{(2k+1)\pi i}{d}}$, $r > 0$ and $k = 0, 1, \ldots, d-1$. By Lemma 5.1, the semi-lines are forward invariant under $T_{p_d,1}$. In fact, we have that $T_{p_d,1}(re^{\frac{(2k+1)\pi i}{d}}) = e^{\frac{(2k+1)\pi i}{d}} R_d(r)$, where

$$R_d(r) := \frac{d(d+1)r^{2d+1}[(d+1)r^d + 1]^d - d^{d+1}r^{(d+1)^2}}{[(d+1)r^d + 1]^{d+2}}.$$

Then, if we can prove that for every $r > 0$ we have $0 < R_d(r) < r$, we can conclude that $\mathcal{A}^*_{T_{p_d,1}}(0)$ is an unbounded set for every d. In that case, we can also state that $\mathcal{A}^*_{T_{p_d,1}}(0)$ has at least d accesses to infinity. Since the denominator of R_d is always positive for every $r > 0$, the inequality $0 < R_d(r)$ is equivalent to

$$d(d+1)r^{2d+1}[(d+1)r^d + 1]^d - d^{d+1}r^{(d+1)^2} > 0. \tag{1}$$

Expanding the last expression using the Binomial expansion, we obtain that inequality (1) becomes

$$d(d+1)\sum_{j=0}^{d-1}\binom{d}{d-j}(d+1)^j r^{dj+2d+1} + d[(d+1)^{d+1} - d^d]r^{(d+1)^2} > 0.$$

Notice that, since $(d+1)^{d+1} - d^d > 0$ for every positive integer d, we obtain that the inequality holds for every $r > 0$.

The inequality $R_d(r) < r$ can be written as $S_d(r) < 0$, where S_d is defined as

$$S_d(r) := d(d+1)r^{2d+1}[(d+1)r^d + 1]^d - d^{d+1}r^{(d+1)^2} - r[(d+1)r^d + 1]^{d+2}.$$

Using the Binomial expansion, we can rewrite the last expression:

$$S_d(r) = d(d+1)\sum_{j=0}^{d-1}\binom{d}{d-j}(d+1)^j r^{dj+2d+1} + d[(d+1)^{d+1} - d^d]r^{(d+1)^2}$$
$$- \sum_{j=-2}^{d}\binom{d+2}{d-j}(d+1)^{j+2} r^{dj+2d+1}.$$

Now, arranging terms,

$$S_d(r) = [d((d+1)^{d+1} - d^d) - (d+1)^{d+2}]r^{(d+1)^2} - \sum_{j=-2}^{0}\binom{d+2}{d-j}(d+1)^{j+2} r^{dj+2d+1}$$
$$+ \sum_{j=0}^{d-1}\left[d(d+1)\binom{d}{d-j}(d+1)^j - \binom{d+2}{d-j}(d+1)^{j+2}\right] r^{dj+2d+1}.$$

Observe that $d((d+1)^{d+1} - d^d) - (d+1)^{d+2} = -(d+1)^{d+1} - d^{d+1} < 0$ and

$$d(d+1)\binom{d}{d-j}(d+1)^j - \binom{d+2}{d-j}(d+1)^{j+2} = (d+1)^{j+1}\left[\frac{d!d - (d+2)!(d+1)}{(d-j)!j!}\right] < 0.$$

Hence, all the coefficients of the polynomial S_d are negative. Therefore, we can conclude that for $r > 0$, $S_d(r) < r$, which completes the proof. □

It still needs to be proven that the immediate basins of attraction of the dth roots of unity are unbounded. This is a more challenging part of the proof, as attempting to apply the same arguments used earlier leads to difficulties in establishing bounds for the expressions. However, a recent study confirms that this holds for all integers $d \geq 2$. In fact, the case of Traub's method applied to the family $p(z) = z(z^d - 1)$ has already been fully resolved [4].

6. Conclusions

With this paper, we are contributing towards demonstrating that the immediate basins of attraction of the damped Traub's method are unbounded and simply connected sets. We have been able to prove with complete generality the unboundedness of the method when $\delta \approx 0$ and we analyze a particular case, the family $p_d(z) = z(z^d - 1)$. Our findings indicate that analyzing the topological properties of this method is not a straightforward and that a comprehensive proof requires different approaches from those used in [3].

References

[1] K. Barański, N. Fagella, X. Jarque, B. Karpińska, Connectivity of Julia sets of Newton maps: a unified approach, *Rev. Mat. Iberoam.* **34(3)** (2018), 1211–1228.

[2] A.F. Beardon, *Iteration of Rational Functions. Complex Analytic Dynamical Systems*, Grad. Texts in Math. **132**, Springer-Verlag, New York, 1991.

[3] J. Canela, V. Evdoridou, A. Garijo, X. Jarque, On the basins of attraction of a one-dimensional family of root finding algorithms: from Newton to Traub, *Math. Z.* **303(3)** (2023), Paper no. 55, 22 pp.

[4] J. Canela, A. Garijo, X. Jarque, Boundedness and simple connectivity of the basins of attraction for some numerical methods, Preprint (2025). `https://arxiv.org/abs/2507.22704`.

[5] L. Carleson, T.W. Gamelin, *Complex Dynamics*, Universitext Tracts Math., Springer-Verlag, New York, 1993.

[6] A. Cordero, A. Ferrero, J.R. Torregrosa, Damped Traub's method: convergence and stability, *Math. Comput. Simulation* **119** (2016), 57–68.

[7] R.L. Devaney, Singular perturbations of complex polynomials, *Bull. Amer. Math. Soc. (N.S.)* **50(3)** (2013), 391–429.

[8] J. Hubbard, D. Schleicher, S. Sutherland, How to find all roots of complex polynomials by Newton's method, *Invent. Math.* **146(1)** (2001), 1–33.

[9] J. Milnor, *Dynamics in One Complex Variable*, Third edition, Ann. of Math. Stud. **160**, Princeton University Press, Princeton, NJ, 2006.

[10] F. Przytycki, Remarks on the simple connectedness of basins of sinks for iterations of rational maps, in: *Dynamical Systems and Ergodic Theory*, Banach Center Publ. **23**, PWN—Polish Scientific Publishers, Warsaw, 1989, pp. 229–235.

[11] M. Shishikura, The connectivity of the Julia set and fixed points, in: *Complex Dynamics*, A K Peters, Ltd., Wellesley, MA, 2009, pp. 257–276.

[12] J.E. Vázquez-Lozano, A. Cordero, J.R. Torregrosa, Dynamical analysis on cubic polynomials of damped Traub's method for approximating multiple roots, *Appl. Math. Comput.* **328** (2018), 82–99.

REPORTS@SCM
AN ELECTRONIC JOURNAL OF THE
SOCIETAT CATALANA DE MATEMÀTIQUES

Convergence of generalized MIT bag models

Joaquim Duran i Lamiel
Centre de Recerca Matemàtica
jduran@crm.cat

Resum *(CAT)*
Estudiem propietats espectrals dels models de bossa de l'MIT generalitzats. Aquests són operadors de Dirac $\{\mathcal{H}_\tau\}_{\tau\in\mathbb{R}}$ actuant en dominis de $\mathbb{R}^3$ amb condicions de frontera que generen confinament. Estudiant la convergència en el sentit de la resolvent dels operadors $\mathcal{H}_\tau$ cap als operadors límit $\mathcal{H}_{\pm\infty}$ quan $\tau\to\pm\infty$, provem que certes propietats espectrals s'hereden al llarg de la parametrització. Aquests resultats, obtinguts parcialment al treball de fi de màster [3], són nous i s'han publicat a [4].

Abstract *(ENG)*
We study spectral properties of generalized MIT bag models. These are Dirac operators $\{\mathcal{H}_\tau\}_{\tau\in\mathbb{R}}$ acting on domains of $\mathbb{R}^3$ with confining boundary conditions. By studying the resolvent convergence of the operators $\mathcal{H}_\tau$ towards the limiting operators $\mathcal{H}_{\pm\infty}$ as $\tau\to\pm\infty$, we prove that certain spectral properties are inherited throughout the parametrization. These results, partially obtained in the master's thesis [3], are new and have been published in [4].

Societat Catalana de Matemàtiques
Institut d'Estudis Catalans

Keywords: *Dirac operator, spectral theory, MIT bag model, shape optimization, resolvent convergence.*
MSC (2020): *Primary 35P05, 35Q40. Secondary 47A10, 81Q10.*

Received: *June 18, 2025.*
Accepted: *August 4, 2025.*

1. Introduction

The equation that governs all relativistic quantum processes is called *Dirac equation*. In $\mathbb{R}^3$, it is a system of four complex valued linear PDEs of first order in time and space variables. For a spin-1/2 free particle of mass $m \geq 0$, one can write the Dirac equation in matricial form as

$$i\frac{\partial}{\partial t}\psi(x,t) = (-i\alpha\cdot\nabla + m\beta)\psi(x,t), \quad x \in \mathbb{R}^3,\ t \geq 0, \tag{1}$$

where $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ and β are the so-called *Dirac matrices*,

$$\alpha_j := \begin{pmatrix} 0 & \sigma_j \\ \sigma_j & 0 \end{pmatrix} \text{ for } j = 1,2,3, \quad \text{and} \quad \beta := \begin{pmatrix} I_2 & 0 \\ 0 & -I_2 \end{pmatrix} \quad \text{with} \quad I_2 := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

given by the *Pauli matrices*

$$\sigma_1 := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 := \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

and where

$$\psi(x,t) = \begin{pmatrix} \psi_1(x,t) \\ \psi_2(x,t) \\ \psi_3(x,t) \\ \psi_4(x,t) \end{pmatrix} \in \mathbb{C}^4$$

is the so-called *wave function* of the particle. Here, $\nabla = (\partial_1, \partial_2, \partial_3)$ denotes the gradient in $\mathbb{R}^3$, and as customary we use the notation $\alpha\cdot\nabla = \alpha_1\partial_1 + \alpha_2\partial_2 + \alpha_3\partial_3$. In Cartesian coordinates, the differential operator in the right-hand side of (1) writes as

$$-i\alpha\cdot\nabla + m\beta = \begin{pmatrix} m & 0 & -i\partial_3 & -i\partial_1 - \partial_2 \\ 0 & m & -i\partial_1 + \partial_2 & i\partial_3 \\ -i\partial_3 & -i\partial_1 - \partial_2 & -m & 0 \\ -i\partial_1 + \partial_2 & i\partial_3 & 0 & -m \end{pmatrix}.$$

Notice that if one diagonalizes this operator (taking into account boundary conditions), one can solve the time-dependent Dirac equation (1) using the method of separation of variables. Hence, the time-dependent problem reduces to a stationary eigenvalue problem of the form

$$\begin{cases} (-i\alpha\cdot\nabla + m\beta)\varphi = \lambda\varphi & \text{in } \Omega, \\ \text{boundary conditions for } \varphi & \text{on } \partial\Omega, \end{cases}$$

where $\Omega \subseteq \mathbb{R}^3$ is the domain where the particle evolves, $\varphi\colon \Omega \to \mathbb{C}^4$, and the boundary conditions typically depend on physical constraints. The eigenvalues λ provide relevant information to understand the evolution of the system, hence this motivates their study and understanding. This is what we do in this work, for some prescribed boundary conditions.

2. Generalized MIT bag models

Dirac operators acting on domains $\Omega \subset \mathbb{R}^3$ with C^2 boundary are used in relativistic quantum mechanics to describe particles that are confined in a box. The so-called *MIT bag model* is a very remarkable example, which was introduced in the 1970s as a simplified model to study confinement of quarks in hadrons (like quarks up and down inside a proton). It is the operator $\mathcal{H}_0$ defined by

$$\begin{aligned} \mathrm{Dom}(\mathcal{H}_0) &:= \{\varphi \in H^1(\Omega) \otimes \mathbb{C}^4 : \varphi = -i\beta(\alpha \cdot \nu)\varphi \text{ on } \partial\Omega\}, \\ \mathcal{H}_0\varphi &:= (-i\alpha \cdot \nabla + m\beta)\varphi \quad \text{for all } \varphi \in \mathrm{Dom}(\mathcal{H}_0). \end{aligned}$$

Here, ν denotes the outward unit normal vector on $\partial\Omega$, and $H^1(\Omega)$ is the standard Sobolev space of first weak derivatives in $L^2(\Omega)$, namely

$$H^1(\Omega) := \{f \in L^2(\Omega) : \|f\|_{H^1(\Omega)} < \infty\}, \quad \text{where} \quad \|f\|_{H^1(\Omega)} := (\|f\|^2_{L^2(\Omega)} + \|\nabla f\|^2_{L^2(\Omega)})^{1/2}.$$

For the sake of notation, in the sequel we shall denote $H^1(\Omega) \otimes \mathbb{C}^4$ as $H^1(\Omega)^4$, and similarly $L^2(\Omega) \otimes \mathbb{C}^4$ as $L^2(\Omega)^4$.

Motivated by some physical considerations, in [1] it was studied the family of Dirac operators with confining boundary conditions defined for $\tau \in \mathbb{R}$ by

$$\begin{aligned} \mathrm{Dom}(\mathcal{H}_\tau) &:= \{\varphi \in H^1(\Omega)^4 : \varphi = i(\sinh\tau - \cosh\tau\,\beta)(\alpha \cdot \nu)\varphi \text{ on } \partial\Omega\}, \\ \mathcal{H}_\tau\varphi &:= (-i\alpha \cdot \nabla + m\beta)\varphi \quad \text{for all } \varphi \in \mathrm{Dom}(\mathcal{H}_\tau). \end{aligned} \tag{2}$$

Notice that the MIT bag model corresponds to $\tau = 0$ —this was the main reason in [1] to call the operators $\mathcal{H}_\tau$ in (2) *generalized MIT bag models*. For $\tau \in \mathbb{R}$, the operator $\mathcal{H}_\tau$ is self-adjoint in $L^2(\Omega)^4$ by [2, Proposition 5.15]. Moreover, from [1, Lemma 1.2] we know that its spectrum $\sigma(\mathcal{H}_\tau)$ is contained in $\mathbb{R} \setminus [-m, m]$ and is purely discrete. In particular, the essential spectrum $\sigma_{\mathrm{ess}}(\mathcal{H}_\tau)$ is empty for all $\tau \in \mathbb{R}$. Furthermore, $\lambda \in \sigma(\mathcal{H}_\tau)$ if and only if $-\lambda \in \sigma(\mathcal{H}_{-\tau})$. Thanks to this odd symmetry, one can reduce the study of the spectral properties of the generalized MIT bag models to the study of $\sigma(\mathcal{H}_\tau) \cap (m, +\infty)$ for $\tau \in \mathbb{R}$.

A spectral study of the mapping $\tau \mapsto \mathcal{H}_\tau$ was carried out in [1], where the following result was shown. In its statement, $-\Delta_D$ denotes the self-adjoint realization of the Dirichlet Laplacian in $L^2(\Omega)$, and $\sigma(-\Delta_D)$ denotes its spectrum.

Theorem 2.1 ([1, Theorem 1.4])**.** *The eigenvalues of $\mathcal{H}_\tau$ can be parametrized by increasing real analytic functions of τ. Moreover, if $\tau \mapsto \lambda(\tau) \in \sigma(\mathcal{H}_\tau) \cap (m, +\infty)$ is a continuous function defined on an interval $I \subset \mathbb{R}$, then the following holds:*

(i) If $I = (-\infty, \tau_0)$ for some $\tau_0 \in \mathbb{R}$, then $\lambda(-\infty) := \lim_{\tau\downarrow-\infty} \lambda(\tau)$ exists and belongs to $[m, +\infty)$. In addition,

$$\lambda(-\infty) = \begin{cases} m \text{ if } \lambda(\tau) \leq \sqrt{\min\sigma(-\Delta_D) + m^2} \text{ for some } \tau \in I, \\ \sqrt{\lambda_D + m^2} \text{ for some } \lambda_D \in \sigma(-\Delta_D) \text{ otherwise.} \end{cases}$$

(ii) If $I = (\tau_0, +\infty)$ for some $\tau_0 \in \mathbb{R}$, then $\lambda(+\infty) := \lim_{\tau\uparrow+\infty} \lambda(\tau)$ exists as an element of the set $(m, +\infty]$. In addition, if $\lambda(+\infty) < +\infty$, then

$$\lambda(+\infty) = \sqrt{\lambda_D + m^2} \quad \text{for some } \lambda_D \in \sigma(-\Delta_D).$$

This result establishes a clear connection between the spectrum of the Dirac operator $\mathcal{H}_\tau$ as $\tau \to \pm\infty$ and the spectrum of the Dirichlet Laplacian $-\Delta_D$. In [1, Remark 4.4] it was left as an open question to investigate which should be the limiting operators of $\mathcal{H}_\tau$ as $\tau \to \pm\infty$, and in which sense the convergence holds true. The answer was developed in the master's thesis [3] and then published in [4]. In the present work, we review the results obtained.

3. Convergence as τ moves in $\mathbb{R}$

In order to guess who the limiting operators might be, we first make an observation. Writing $\varphi \in \mathrm{Dom}(\mathcal{H}_\tau)$ in components[1] as $\varphi = (u, v)^\intercal$, the boundary condition

$$\varphi = i(\sinh \tau - \cosh \tau \, \beta)(\alpha \cdot \nu)\varphi$$

rewrites as $u = -ie^{-\tau}(\sigma \cdot \nu)v$. Formally, this equation forces u and v to vanish on $\partial\Omega$ in the limits $\tau \uparrow +\infty$ and $\tau \downarrow -\infty$, respectively. This leads to consider the so-called *Dirac operators with zigzag type boundary conditions* studied in [6], which are defined by

$$\begin{aligned} \mathrm{Dom}(\mathcal{H}_{+\infty}) &:= \{\varphi = (u, v)^\intercal : u \in H^1_0(\Omega)^2,\ v \in L^2(\Omega)^2,\ \alpha \cdot \nabla\varphi \in L^2(\Omega)^4\}, \\ \mathcal{H}_{+\infty}\varphi &:= (-i\alpha \cdot \nabla + m\beta)\varphi \quad \text{for all } \varphi \in \mathrm{Dom}(\mathcal{H}_{+\infty}) \end{aligned} \tag{3}$$

—here $H^1_0(\Omega)^2$ is the subspace of functions in $H^1(\Omega)^2$ with zero trace—, and

$$\begin{aligned} \mathrm{Dom}(\mathcal{H}_{-\infty}) &:= \{\varphi = (u, v)^\intercal : u \in L^2(\Omega)^2,\ v \in H^1_0(\Omega)^2,\ \alpha \cdot \nabla\varphi \in L^2(\Omega)^4\}, \\ \mathcal{H}_{-\infty}\varphi &:= (-i\alpha \cdot \nabla + m\beta)\varphi \quad \text{for all } \varphi \in \mathrm{Dom}(\mathcal{H}_{+\infty}). \end{aligned} \tag{4}$$

From [6, Theorem 1.1 and Lemma 3.2] we know that $\mathcal{H}_{\pm\infty}$ are self-adjoint in $L^2(\Omega)^4$ and that their spectra are characterized by the spectrum of the Dirichlet Laplacian. More specifically,

$$\begin{aligned} \sigma(\mathcal{H}_{+\infty}) &= \{-m\} \cup \{\pm\sqrt{\lambda_D + m^2} : \lambda_D \in \sigma(-\Delta_D)\}, \\ \sigma(\mathcal{H}_{-\infty}) &= \{m\} \cup \{\pm\sqrt{\lambda_D + m^2} : \lambda_D \in \sigma(-\Delta_D)\}, \end{aligned} \tag{5}$$

and $\mp m \in \sigma_{\mathrm{ess}}(\mathcal{H}_{\pm\infty})$ is an eigenvalue of infinite multiplicity.

Observe that the description (5) of $\sigma(\mathcal{H}_{\pm\infty})$ is in agreement with the limiting spectrum stated in Theorem 2.1. This heuristically motivates to propose the operators $\mathcal{H}_{\pm\infty}$ defined in (3) and (4) as the limiting operators of $\mathcal{H}_\tau$, as $\tau \to \pm\infty$. To see in which sense the convergence holds true, we study the resolvent convergence of $\mathcal{H}_\tau$ to $\mathcal{H}_{\pm\infty}$ as $\tau \to \pm\infty$; see [8, Chapter 8] for a survey on resolvent convergence.

Theorem 3.1 ([4, Theorem 1.2]). *Given $\tau \in \mathbb{R}$, let $\mathcal{H}_\tau$ be the operator defined in* (2). *Let $\mathcal{H}_{+\infty}$ and $\mathcal{H}_{-\infty}$ be the operators defined in* (3) *and* (4), *respectively. Then, $\mathcal{H}_\tau$ converges to $\mathcal{H}_{\pm\infty}$ in the strong resolvent sense as $\tau \to \pm\infty$. That is, for every $f \in L^2(\Omega)^4$*

$$\lim_{\tau \to \pm\infty} \|((\mathcal{H}_{\pm\infty} - \lambda)^{-1} - (\mathcal{H}_\tau - \lambda)^{-1})f\|_{L^2(\Omega)^4} = 0 \quad \text{for all } \lambda \in \mathbb{C} \setminus \mathbb{R}. \tag{6}$$

[1] The notation $\varphi = (u, v)^\intercal$ refers to the decomposition of $\varphi \colon \Omega \to \mathbb{C}^4$ in upper and lower components, that is, if $\varphi = (\varphi_1, \varphi_2, \varphi_3, \varphi_4)^\intercal$ with $\varphi_j \colon \Omega \to \mathbb{C}$ for $j = 1, 2, 3, 4$, then $u = (\varphi_1, \varphi_2)^\intercal$ and $v = (\varphi_3, \varphi_4)^\intercal$.

A proof of this theorem based on directly estimating the difference of resolvents in (6) can be found in [3, Section 3.2], and an alternative proof based on the notion of strong graph limit [8, Definition in p. 293] can be found both in [3, Section 3.1] and in [4, Section 2]. An immediate consequence of this theorem is the following result, which is an improvement of item (ii) in Theorem 2.1 for the first positive eigenvalue of $\mathcal{H}_\tau$.

Corollary 3.2 ([4, Corollary 1.3])**.** *For every $\tau \in \mathbb{R}$, denote the first positive eigenvalue of $\mathcal{H}_\tau$ in Ω by $\lambda_\Omega(\tau) := \min(\sigma(\mathcal{H}_\tau) \cap (m, +\infty))$. Then, $\lim_{\tau\uparrow+\infty} \lambda_\Omega(\tau) = \sqrt{\Lambda_\Omega + m^2}$, where $\Lambda_\Omega := \min \sigma(-\Delta_D)$ is the first eigenvalue of the Dirichlet Laplacian in Ω.*

It is remarkable to point out that Theorem 3.1 does not ensure that the convergence in (6) is uniform in the unit ball of $L^2(\Omega)^4$, but only pointwise for every $f \in L^2(\Omega)^4$. Actually, we now justify that the convergence can not be uniform —in the language of resolvents, this means that $\mathcal{H}_\tau$ can not converge to $\mathcal{H}_{\pm\infty}$ in the norm resolvent sense as $\tau \to \pm\infty$; see [8, Definition in p. 284] or Theorem 3.4 below—: indeed, if there was convergence in the norm resolvent sense, [9, Satz 9.24] would lead to $\lim_{\tau\to\pm\infty} \sigma_{\mathrm{ess}}(\mathcal{H}_\tau) = \sigma_{\mathrm{ess}}(\mathcal{H}_{\pm\infty})$, but this is impossible since $\sigma_{\mathrm{ess}}(\mathcal{H}_{\pm\infty}) \neq \emptyset$ —recall that $\mp m$ is an eigenvalue of infinite multiplicity— and $\sigma_{\mathrm{ess}}(\mathcal{H}_\tau) = \emptyset$ for all $\tau \in \mathbb{R}$ —because $\sigma(\mathcal{H}_\tau)$ is purely discrete.

This argument shows that the essential eigenvalue $\mp m \in \sigma_{\mathrm{ess}}(\mathcal{H}_{\pm\infty})$ prevents $\mathcal{H}_\tau$ from converging to $\mathcal{H}_{\pm\infty}$ in the norm resolvent sense. It is then natural to ask whether the norm resolvent convergence could be achieved if, in some sense, the study was restricted to $\sigma(\mathcal{H}_{\pm\infty}) \setminus \{\mp m\}$. An affirmative answer holds true in the following sense. Denote

$$\ker(\mathcal{H}_{\pm\infty} \pm m) := \{\psi \in \mathrm{Dom}(\mathcal{H}_{\pm\infty}) \subset L^2(\Omega)^4 : (\mathcal{H}_{\pm\infty} \pm m)\psi = 0\},$$
$$\ker(\mathcal{H}_{\pm\infty} \pm m)^\perp := \{\varphi \in L^2(\Omega)^4 : \langle \varphi, \psi \rangle_{L^2(\Omega)^4} = 0 \text{ for all } \psi \in \ker(\mathcal{H}_{\pm\infty} \pm m)\}.$$

Since $\ker(\mathcal{H}_{\pm\infty} \pm m)^\perp$ is a closed subspace of $L^2(\Omega)^4$, the orthogonal projection

$$P_\pm \colon L^2(\Omega)^4 \to \ker(\mathcal{H}_{\pm\infty} \pm m)^\perp \subset L^2(\Omega)^4 \tag{7}$$

is a well-defined bounded self-adjoint operator in $L^2(\Omega)^4$. Moreover, from (5) we know that $\ker(\mathcal{H}_{\pm\infty} \pm m)^\perp \neq \{0\}$ and, thus, $\|P_\pm\|_{L^2(\Omega)^4\to L^2(\Omega)^4} = 1$.

Theorem 3.3 ([4, Theorem 1.4])**.** *Given $\tau \in \mathbb{R}$, let $\mathcal{H}_\tau$ be the operator defined in (2). Let $\mathcal{H}_{+\infty}$ and $\mathcal{H}_{-\infty}$ be the operators defined in (3) and (4), respectively. Then,*

$$\lim_{\tau\to\pm\infty} \|P_\pm((\mathcal{H}_{\pm\infty} - \lambda)^{-1} - (\mathcal{H}_\tau - \lambda)^{-1})\|_{L^2(\Omega)^4\to L^2(\Omega)^4} = 0 \quad \text{for all } \lambda \in \mathbb{C} \setminus \mathbb{R},$$

where $P_\pm$ are the orthogonal projections defined in (7).

A proof of this theorem can be found in [4, Section 3]. As we mentioned after Corollary 3.2, the difference of resolvents $(\mathcal{H}_{\pm\infty} - \lambda)^{-1} - (\mathcal{H}_\tau - \lambda)^{-1}$ does not converge to zero in operator norm as $\tau \to \pm\infty$. However, if we write this difference as

$$(\mathcal{H}_{\pm\infty} - \lambda)^{-1} - (\mathcal{H}_\tau - \lambda)^{-1} = (P_\pm + (1 - P_\pm))((\mathcal{H}_{\pm\infty} - \lambda)^{-1} - (\mathcal{H}_\tau - \lambda)^{-1}),$$

then Theorem 3.3 shows that the eigenvalue $\mp m$ is indeed the only obstruction for having norm resolvent convergence of $\mathcal{H}_\tau$ to $\mathcal{H}_{\pm\infty}$ as $\tau \to \pm\infty$, since $(1 - P_\pm)(L^2(\Omega)^4) = \ker(\mathcal{H}_{\pm\infty} \pm m)$.

Although the main interest is the study of the convergence of $\mathcal{H}_\tau$ in a resolvent sense as $\tau \to \pm\infty$, for the sake of completeness we also study the convergence when τ approaches a finite value $\tau_0 \in \mathbb{R}$.

Theorem 3.4. *Given $\tau \in \mathbb{R}$, let $\mathcal{H}_\tau$ be the operator defined in (2). Then, for every $\tau_0 \in \mathbb{R}$, $\mathcal{H}_\tau$ converges to $\mathcal{H}_{\tau_0}$ in the norm resolvent sense as $\tau \to \tau_0$. That is,*

$$\lim_{\tau\to\pm\infty} \|(\mathcal{H}_{\pm\infty} - \lambda)^{-1} - (\mathcal{H}_\tau - \lambda)^{-1}\|_{L^2(\Omega)^4\to L^2(\Omega)^4} = 0 \quad \textit{for all } \lambda \in \mathbb{C} \setminus \mathbb{R}.$$

A proof of this theorem based on the fact that the resolvent operator $(\mathcal{H}_\tau - \lambda)^{-1}$ is real analytic in τ in a neighborhood of τ_0 —given by [1, Lemma 3.1]— can be found in [3, Section 3.4]. An alternative proof based on estimating the operator norm of the difference of resolvents can be found in [4, Section 4].

4. Shape optimization

A hot open problem in spectral geometry is to prove that the first positive eigenvalue $\lambda_\Omega(\tau)$ of $\mathcal{H}_\tau$ is minimal, among all bounded C^2 domains $\Omega \subset \mathbb{R}^3$ with prescribed volume, when Ω is a ball; see [1, Conjecture 1.8]. The analogous statement for the first eigenvalue of the Dirichlet Laplacian, $\Lambda_\Omega := \min \sigma(-\Delta_D)$, is known to be true, and it is the so-called *Faber–Krahn inequality* —proven independently by Faber in 1923 and Krahn in 1925 [5, 7], asserting that $\Lambda_\Omega > \Lambda_B$ whenever $\Omega \subset \mathbb{R}^3$ is a bounded domain with Lipschitz boundary different from a ball B with the same volume.

As an application of the results obtained in [3, 4] and presented in this paper, we conclude with a statement supporting (but not proving) the optimality of the ball for $\lambda_\Omega(\tau)$. On the one hand, $\tau \mapsto \lambda_\Omega(\tau)$ is an increasing and continuous function in $\mathbb{R}$, that converges to m as $\tau \to -\infty$ —by Theorem 2.1— and that converges to $\sqrt{\Lambda_\Omega + m^2}$ as $\tau \uparrow +\infty$, by Corollary 3.2; in particular, $\tau \mapsto \lambda_\Omega(\tau)$ is bijective from $\mathbb{R}$ to $(m, \sqrt{\Lambda_\Omega + m^2})$. On the other hand, if Ω is not a ball, then by the Faber–Krahn inequality we have $(m, \sqrt{\Lambda_B + m^2}) \subsetneq (m, \sqrt{\Lambda_\Omega + m^2})$. Therefore, there exists a large enough $\tau_\Omega \in \mathbb{R}$ such that $\lambda_\Omega(\tau) \in (\sqrt{\Lambda_B + m^2}, \sqrt{\Lambda_\Omega + m^2})$ for all $\tau \geq \tau_\Omega$. Since $\lambda_B(\tau) < \sqrt{\Lambda_B + m^2}$ for all such τ —by Theorem 2.1 and Corollary 3.2—, we get the following.

Proposition 4.1. *Let $\Omega \subset \mathbb{R}^3$ be a bounded domain with C^2 boundary, and let B be a ball such that $|\Omega| = |B|$. If Ω is not a ball, then there exists $\tau_\Omega \in \mathbb{R}$ such that $\lambda_B(\tau) < \lambda_\Omega(\tau)$ for all $\tau \geq \tau_\Omega$.*

It is very remarkable to say that the large enough $\tau_\Omega \in \mathbb{R}$ ensuring the optimality of the ball for the first positive eigenvalue $\lambda_\Omega(\tau)$ in the regime $\tau \geq \tau_\Omega$ depends itself on Ω. Hence, from Proposition 4.1 one can *not* ensure that there exists a large enough $\tau_\star \in \mathbb{R}$ for which $\lambda_\Omega(\tau) > \lambda_B(\tau)$ for all $\tau \geq \tau_\star$ and *every* bounded C^2 domain Ω different from a ball B with the same volume. To prove or disprove the existence of such $\tau_\star$ also remains as an open and challenging problem.

Acknowledgements

The author is supported by the Spanish grants PID2021-123903NB-I00 and RED2022-134784-T funded by MCIN/AEI/10.13039/501100011033, by ERDF "A way of making Europe", and by the Catalan grant 2021-SGR-00087. This work is supported by the Spanish State Research Agency, through the Severo Ochoa and María de Maeztu Program for Centers and Units of Excellence in R&D (CEX2020-001084-M), and more specifically by the grant CEX2020-001084-M-20-1. The author acknowledges CERCA Programme/Generalitat de Catalunya for institutional support.

The author thanks Societat Catalana de Matemàtiques and Institut d'Estudis Catalans for awarding with a runner up the master's thesis [3] in the 62nd edition of the Évariste Galois Award. The author also thanks Albert Mas for his guidance, dedication, and commitment.

References

[1] N. Arrizabalaga, A. Mas, T. Sanz-Perela, L. Vega, Eigenvalue curves for generalized MIT bag models, *Comm. Math. Phys.* **397(1)** (2023), 337–392.

[2] J. Behrndt, M. Holzmann, A. Mas, Self-adjoint Dirac operators on domains in $\mathbb{R}^3$, *Ann. Henri Poincaré* **21(8)** (2020), 2681–2735.

[3] J. Duran, Spectral gap of generalized MIT bag models, Master's Thesis, Universitat Politècnica de Catalunya, 2024. `https://upcommons.upc.edu/handle/2117/400748?locale-attribute=en`.

[4] J. Duran, A. Mas, Convergence of generalized MIT bag models to Dirac operators with zigzag boundary conditions, *Anal. Math. Phys.* **14(4)** (2024), Paper no. 85, 23 pp.

[5] C. Faber, *Beweiss, dass unter allen homogenen Membrane von gleicher Fläche und gleicher Spannung die kreisförmige die tiefsten Grundton gibt*, Sitzungsber.-Bayer. Akad. Wiss., Math.-Phys. Munich. (1923), 169–172.

[6] M. Holzmann, A note on the three dimensional dirac operator with zigzag type boundary conditions, *Complex Anal. Oper. Theory* **15(3)** (2021), Paper no. 47, 15 pp.

[7] E. Krahn, Über eine von Rayleigh formulierte Minimaleigenschaft des Kreises, *Math. Ann.* **94** (1925), 97–100.

[8] M. Reed, B. Simon, *Methods of Modern Mathematical Physics. I. Functional Analysis*, Second edition, Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York, 1980.

[9] J. Weidmann, *Lineare Operatoren in Hilberträumen. Teil 1*, Grundlagen, Mathematische Leitfäden, B. G. Teubner, Stuttgart, 2000.

REPORTS@SCM

AN ELECTRONIC JOURNAL OF THE
SOCIETAT CATALANA DE MATEMÀTIQUES

Monodromy conjecture for Newton non-degenerate hypersurfaces

Oriol Baeza Guasch
Universitat Politècnica de Catalunya
oriol.baeza@estudiantat.upc.edu

Resum *(CAT)*
Aquest treball estudia la Conjectura Forta de la Monodromia (SMC) en la versió topològica. Després d'introduir els conceptes de resolució de singularitats, polinomi de Bernstein–Sato i la funció zeta, esbocem els resultats involucrats en la demostració de la SMC per a singularitats Newton no degenerades (NND). Aquesta prova requereix, però, hipòtesis addicionals sobre els nombres del residu, i construïm exemples que mostren que no poden ometre's, la qual cosa suggereix que calen altres tècniques per a atacar el cas general.

Abstract *(ENG)*
This work studies the Strong Monodromy Conjecture (SMC) in its topological setting. After introducing the concepts of resolution of singularities, Bernstein–Sato polynomial, and the zeta function, we sketch the results involved in the proof of the SMC for Newton non-degenerate (NND) singularities. This approach requires nonetheless additional hypothesis on the residue numbers, and we construct examples showing that they can't be dropped, which suggests that new techniques are needed to attack the general case.

Societat Catalana de Matemàtiques

Institut d'Estudis Catalans

Keywords: *monodromy, Bernstein–Sato polynomial, resolution of singularities, plane curves, Newton non-degenerate.*
MSC (2020): *Primary 14B05, 32S40. Secondary 14H20, 14J17, 34M35.*

Received: *July 1, 2025.*
Accepted: *July 18, 2025.*

1. Introduction

The monodromy conjecture is a problem in the field of singularity theory in algebraic geometry, formulated by the Japanese mathematician Igusa in the seventies, which relates two invariants of a singularity. On the one hand, the roots of a polynomial arising from a functional equation satisfied by the singularity (the so called Bernstein–Sato polynomial). On the other hand, the poles of the zeta function (in our setting, the topological version), which contains information about a resolution of the singularity. More precisely, the conjecture predicts that every pole of this zeta function is a root of the Bernstein–Sato.

Although the general case remains open, a positive result has been proven for some special cases. In particular, it is known to be true for plane curves (Loeser '88), for Newton non-degenerate (NND) polynomials modulo an hypothesis on the so called *residue numbers* (Loeser '90), as well as in certain hyperplane arrangements, or also semi-quasihomogeneous singularities.

Both in the cases of plane curves and NND polynomials, a more combinatorial approach is possible, which simplifies some computations and allows to use some technical cohomological results. Nonetheless, for the NND case, this comes with the price of adding two hypothesis on the residue numbers. We discuss the possibility of removing the hypothesis, and show that they do not hold in general. Even more, we will see that divisors not satisfying them can still contribute to the poles of the topological zeta function, suggesting that this approach won't work for the general case.

2. Preliminaries

2.1. Complex zeta function and resolution of singularities

Before stating the conjecture, we must introduce the two main objects of the problem: the Bernstein–Sato polynomial and the topological zeta function. Even more, to give some context and motivation of the statement, we must first begin with the complex zeta function.

The complex zeta function, for a polynomial f and a test function ϕ (meaning a complex function $\mathbb{C}^\infty$ with compact support), is defined as

$$Z(s) = Z(f, \phi; s) := \int_{\mathbb{R}^n} |f(x)|^s \phi(x)\, \mathrm{d}x,$$

where technically we must understand this a distribution in the space of test functions. It can be checked that $Z(s)$ converges and is holomorphic in the semiplane $\Re(s) > 0$. Its meromorphic continuation and the distribution of the possible poles was posed as a problem by I. Gelfand [11, §3.I], and solved in two different manners.

On one hand, we can use a resolution of singularities (guaranteed in characteristic 0 by Hironaka [12]). Recall that an embedded resolution of a polynomial f is a proper morphism $\pi\colon Y \to X$ such that Y is smooth, the restriction of π outside the singular locus is an isomorphism, and that around each point in the preimage we have a neighborhood and a chart over which $\pi^* f = u(y) y_{i_1}^{N_1} \cdots y_{i_r}^{N_r}$ with $u(0) \neq 0$ a unit and $N_i \geq 0$ integers. From the local expression, we can write the pullback divisor globally as

$$\operatorname{div}(\pi^* f) = \sum_{j \in J} N_j E_j,$$

where $(E_i)_{i\in J}$ are the irreducible components of $\pi^{-1}(f^{-1}(0))$, each E_i given in local coordinates by $\{x_i = 0\}$, respectively. Another relevant numerical quantity that will appear are the coefficients in the global expression of the pullback of the standard volume form

$$\operatorname{div}(\pi^*(\mathrm{d}x_1 \wedge \cdots \wedge \mathrm{d}x_n)) = \sum_{j\in J}(k_j - 1)E_j.$$

In this context, we can use the resolution as a change of variables in the integral and deduce that the poles of the complex zeta function are of the form $-\frac{k_j+\nu}{N_j}$, with ν a non-negative integer, and (k_j, N_j) the numerical data associated to the exceptional divisors, introduced above.

2.2. Bernstein–Sato polynomial

On the other hand, we can introduce the Bernstein–Sato polynomial, which is an analytical invariant (and not a topological one) of the singularity. First, consider $R := \mathbb{C}[x_1, \dots, x_n]$ (or more generally the ring of holomorphic functions or even formal power series) and denote $\mathscr{D} := R\langle\partial_1, \dots, \partial_n\rangle$ the Weyl algebra. All elements commute except for the relations $\partial_i x_i - x_i\partial_i = 1$, and so it is easy to show that any element (a priori only a linear differential operator) can be written as a finite sum $P = \sum_{\alpha,\beta} a_{\alpha,\beta}x^\alpha\partial^\beta$. For a more gentle introduction and more details on the properties of the Weyl algebra, we refer to [5]. Next, consider the polynomial ring $\mathscr{D}[s] := \mathscr{D} \otimes_{\mathbb{C}} \mathbb{C}[s]$ with new variable s, and note that the free module $R_f[s]\cdot f^s$ has a natural structure of left $\mathscr{D}[s]$-module given by the product rule. Indeed, every element of the module can be written as $\frac{g}{f^k}\cdot f^s$ for some $g(x, s) \in R[s]$, and the action of the partial derivatives is

$$\partial_i \cdot \left(\frac{g}{f^k}\cdot f^s\right) = \partial_i \cdot \left(\frac{g}{f^k}\right)\cdot f^s + \frac{sg}{f^{k+1}}\cdot\frac{\partial f}{\partial x_i}\cdot f^s.$$

The relevant result is then the existence of solutions to the following functional equation, which was first proven by Bernstein [2] for the case of polynomials, and later by Kashiwara [13] and Björk [3] in the cases of holomorphic functions and formal series, respectively.

Theorem 2.1 ([2])**.** *Let $f \in R$ be a polynomial. Then, there exists a polynomial $P(s) \in \mathscr{D}[s]$ and a polynomial $b_{f,P}(s) \in \mathbb{C}[s]$ such that the relation*

$$P(s)f^{s+1} = b_{f,P}(s)f^s \tag{1}$$

holds formally in the $\mathscr{D}$-module $R_f[s]\cdot f^s$.

The set of polynomials $b_{f,P}$ satisfying such a differential equation as above form an ideal in $\mathbb{C}[s]$, so we can consider its monic generator: the *Bernstein–Sato polynomial* of f denoted by $b_f(s)$.

Example 2.2. For $f = x_1^2 + \cdots + x_n^2$ we have $b_f(s) = (s+1)(s+n/2)$, as taking $P(s)$ to be the Laplacian operator (thought as a constant polynomial in $\mathscr{D}[s]$), we have the relation

$$\left(\frac{\partial^2}{\partial x_1^2} + \cdots + \frac{\partial^2}{\partial x_n^2}\right) f^{s+1} = 4(s+1)\left(s+\frac{n}{2}\right) f^s.$$

Example 2.3. For $f = x^2 + y^3$ in $\mathbb{C}[x, y]$ we have the following relation

$$\left[\frac{1}{12}\frac{\partial}{\partial x}\frac{\partial}{\partial y}y\frac{\partial}{\partial x} + \frac{1}{27}\left(\frac{\partial}{\partial y}\right)^3 + \frac{1}{4}\left(s+\frac{7}{6}\right)\left(\frac{\partial}{\partial x}\right)\right] f^{s+1} = (s+1)\left(s+\frac{5}{6}\right)\left(s+\frac{7}{6}\right) f^s,$$

and it can be proved that $b_f(s) = (s+1)(s+\frac{5}{6})(s+\frac{7}{6})$.

Now, having introduced the functional equation (1), the idea is to use it to integrate by parts and obtain a meromorphic continuation of $Z(s)$ to the whole complex plane, as for any $r \in \mathbb{N}$ we have

$$Z(s) = \frac{1}{b_f(s+r-1)\cdots b_f(s+1)b_f(s)} \int_{\mathbb{R}^n} f(x)^{s+r}\phi_r(x)\,\mathrm{d}x, \quad \Re(s) > -r.$$

In this case, it can be seen that the poles of the zeta function are of the form $\lambda - \nu$ for λ a root of $b_f(s)$ and ν a non-negative integer. By comparing this with the previous candidate quantities for the poles, one arrives at the following result.

Theorem 2.4. *Every root of the Bernstein–Sato polynomial b_f is of the form $-\frac{k_j+\nu}{N_j}$ for some $j \in J$ and ν a non-negative integer.*

2.3. Monodromy conjecture

Altogether, the relation between poles of the complex zeta function and the roots of the Bernstein–Sato is clear. Motivated by this, and after computing some examples, Igusa formulated the conjecture for the poles of the p-adic zeta function, and later stated in the topological and motivic settings too.

In the topological version, we have left to introduce the topological zeta function, first defined by Denef and Loeser in [8], who formalized its definition from a heuristic argument taking a limit of the p-adic zeta function, and showing that the following expression is independent of the choice of a resolution (there is currently no known intrinsic definition).

Definition 2.5 (Topological zeta function)**.** Let $f \in \mathbb{C}[x_1, \dots, x_n]$ be a non-constant polynomial, and choose a resolution $\pi\colon Y \to \mathbb{A}^n_{\mathbb{C}}$ of $\{f = 0\}$. The (global) *topological zeta function* of f is

$$Z_{\mathrm{top}}(f; s) := \sum_{I \subset J} \chi(E_I^\circ) \prod_{i \in I} \frac{1}{k_i + N_i s}, \quad E_I^\circ = \bigcup_{j \in I} E_j \setminus \bigcup_{j \notin I} E_j, \text{ for } I \subset J.$$

Then, it is clear from the expressions that the poles are still related with the roots of the Bernstein–Sato, as they both contain information from a resolution. However, the difficulty lies, on one side, in determining which poles remain in the zeta function after possible cancellation, and in the other, what candidate values in Theorem 2.4 are actually roots. The monodromy conjecture partly answers this, by predicting that every pole of the topological zeta function is a root.

Conjecture 2.6 (Topological monodromy conjecture)**.** *Let $f \in \mathbb{C}[x_1, \dots, x_n]$ be a non-constant polynomial. If s_0 is a pole of $Z_{\mathrm{top}}(f; s)$, then*

(i) (standard) $e^{2\pi\imath\Re(s_0)}$ *is an eigenvalue of the monodromy of $f\colon \mathbb{C}^n \to \mathbb{C}$ at a point of $\{f = 0\}$.*

(ii) (strong) s_0 *is a root of the Bernstein–Sato polynomial $b_f(s)$.*

In this work, we always refer to the strong version, which implies the standard (or weak) one thanks to a result by Malgrange ([20, Proposition 7.1]). So far, the approaches for the solved cases do not provide a clear conceptual idea why the result should be true, apart from the analogy with the complex zeta function (see [22, Remark 2.13]).

A key common element in the proofs of the known cases is the study of periods of integrals (see [18, 19]). These objects allow to relate its asymptotic behavior with the monodromy action (that is, the action of a generator of the fundamental group of a punctured disk around the singularity of the homology groups of a fiber X_t of the Milnor fiber). Even more, in this context Malgrange is able to prove that certain quantities appearing in the asymptotic expressions are roots of $b_f(s)$.

Nonetheless, this approach requires the existence of a non-zero cohomology class in $H_n(X_t, \mathbb{C})$. This result is precisely what Deligne and Mostow prove in the case of plane curves ([7, Proposition 2.14]), and Loeser for the case of Newton non-degenerate ([15, Théorème 3.7]) adapting a result by Esnault and Viehweg ([9]).

Furthermore, in the case of plane curves, the cohomological result of existence only applies to rupture divisors of the resolution. For this reason, Loeser complements it with a combinatorial study of a (minimal) resolution of the singularity, represented in the so called *dual graph*. In this graph, each vertex represents a rupture divisor E_i, $i \in J$, of the resolution, and the edges are added and modified after each *blow up*. In this situation, it is possible to study the quotient k_i/N_i and the residue numbers $\varepsilon(i,j) = k_j - N_j k_i/N_i$ over the divisors via recurrences in the dual graph. In particular, it can be proven that the only possible poles contributing to the topological zeta function arise from the rupture divisors, and so the aforementioned cohomological result is enough.

As for the NND case, in the next sections we will see how the technical hypothesis required for the cohomological result lead to additional conditions on the residue numbers. We will show that the relevant hypothesis is the second one, and we will construct examples that prove that they are not always satisfied. Even worse, we will see that the *bad* divisors violating it can contribute with non-zero residue to a pole of the topological zeta function.

3. Newton non-degenerate polynomials

3.1. Definition and properties

Polynomials which are Newton non-degenerate are sometimes also called non-degenerate with respect to the Newton polytope, or simply non-degenerate. The condition that these polynomials satisfy has a more combinatorial flavor, as it is easier described by considering the objects that we will introduce next.

We consider a polynomial $f(x_1, \dots, x_n) = \sum_{p \in \mathbb{N}^n} a_p x_1^p \cdots x_n^p$ such that $f(0) = 0$. For brevity, we will use multi-index notation when convenient $f(x) = \sum_{p \in \mathbb{N}^n} a_p x^p$, and we define its support to be $\operatorname{supp}(f) = \{p \in \mathbb{N}^n \mid a_p \neq 0\}$.

Definition 3.1 (Newton polyhedron). Let $f = \sum_{p \in \mathbb{N}^n} a_p x^p \in \mathbb{C}[x]$ with $f(0) = 0$. We define the *global Newton polyhedron* $\Gamma_{gl}(f)$ of f as the convex hull of $\operatorname{supp}(f)$. Also, we define the *local Newton polyhedron* $\Gamma(f)$ as the convex hull of the set $\bigcup_{p \in \operatorname{supp}(f)} p + (\mathbb{R}_{\geq 0})^n$.

We will use the term *face* of $\Gamma(f)$ to refer to any convex subset τ that can be obtained by intersecting the Newton diagram with a hyperplane H of $\mathbb{R}^n$ such that $\Gamma(f)$ is contained in one of the half-spaces defined by H. Note that we also consider the total polyhedron as a face.

Definition 3.2 (Non-degenerate)**.** We say that f is Newton non-degenerate at 0 if for any face $\tau \subset \Gamma(f)$, the hypersurface defined by the truncation $f^\tau := \sum_{p \in \tau \cap \mathrm{supp}(f)} a_p x^p$ satisfies that the polynomials $x_i \frac{\partial f^\tau}{\partial x_i}$ for $i = 1, \dots, n$ do not vanish at the same time in $(\mathbb{C} \setminus 0)^n$.

This class of NND polynomials is general enough to be of interest, while also allows a more combinatorial treatment of the problem. We will now see how the polyhedron and its dual fan encode the information of a good resolution of the singularity. For that, we first introduce some notation.

Definition 3.3 (N, k, F)**.** For a vector $a \in (\mathbb{R}^+)^n$, we define the quantities $N(a) := \inf_{x \in \Gamma(f)}\{\langle a, x\rangle\}$ and $k(a) := \sum_{i=1}^n a_i$. Also, define the first meet locus $F(a) := \{x \in \Gamma(f) \mid \langle a, x\rangle = N(a)\}$, which is a proper face of $\Gamma(f)$ if $a \neq 0$, and $F(0)$ recovers the whole diagram.

Definition 3.4 (Dual fan)**.** For τ a face of $\Gamma(f)$, we define the *cone associated* to τ as

$$\Delta_\tau := \{a \in (\mathbb{R}_{>0})^n \mid F(a) = \tau\}/\sim, \quad a \sim a' \text{ iff } F(a) = F(a').$$

The collection of these cones for all faces of the Newton polytope as the *dual fan*.

As an example, see the following Figure 1, where we consider the plane curve defined by $f = x^3 - y^2 + 4xy + 3x^2y$, and construct the Newton polyhedron with labeled faces and corresponding truncations (left), and the associated dual fan (right).

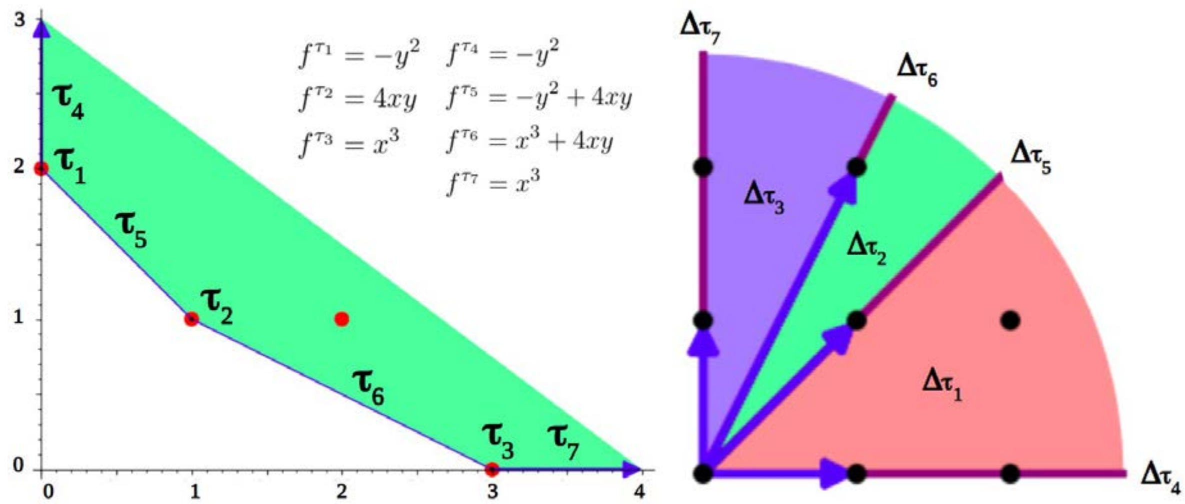

Figure 1: $\Gamma(f)$ and dual fan of the polynomial $f = x^3 - y^2 + 4xy + 3x^2y$.

Next, we recall the following properties of cones.

Definition 3.5 (Cone)**.** A *convex polyhedral cone*, or *cone* for short, is a set

$$C = \{\lambda_1 v_1 + \cdots + \lambda_s v_r \in V \mid \lambda_i \geq 0\} = \mathbb{R}_{\geq 0} v_1 + \cdots + \mathbb{R}_{\geq 0} v_r,$$

where V is an n-dimensional vector space over $\mathbb{R}$, and the vectors $\{v_i\}$ are called the *generators* of the cone. The dimension of C is defined to be the dimension of the smallest vector space containing it.

We say that the cone is *simplicial* if its generating vectors are linearly independent over $\mathbb{R}$. Moreover, we will say it is *simplicial rational* if on top of that the entries of the vectors are integers. We say that the cone is *regular* (or *simple*) if the set of generating vectors can be extended to a base of the $\mathbb{Z}$-module $\mathbb{Z}^n$.

Then, the key theorem from toric geometry that we will need is the following.

Theorem 3.6 ([14, pp. 32–25]). *Let Δ be a cone generated by vectors $v_1, \dots, v_r \in \mathbb{R}^n \setminus \{0\}$. There exists a finite partition of Δ in cones Δ_i, such that each cone is generated by a subset of linearly independent vectors of $\{v_1, \dots, v_r\}$. Moreover, if Δ is simplicial rational, a partition in regular cones can be obtained by introducing suitable new generating rays.*

Such a regular simplicial subdivision can be obtained algorithmically (see also [1, §8.2.2]), and in particular applying it to all the cones in the dual fan, we deduce the next result.

Theorem 3.7 ([1, Lemma 8.7]). *There exists a fan consisting of regular simplicial cones which is obtained as a subdivision of the dual fan associated to the Newton polyhedron.*

Remark 3.8. Notice, however, that we haven't claimed anything about uniqueness, as there can exist multiple valid subdivisions in simplicial regular cones.

3.2. Resolution and topological zeta function from the dual fan

Back to resolution of singularities, we know thanks to the result by Hironaka that we always have one, and that it can be obtained as a composition of blowups. However, in this setup we will obtain a resolution more directly via a single toric blowup, which we define next (see a more detailed introduction in [21] and [17], and also [10, 6] for a complete treatment of toric varieties).

Definition 3.9 (Toric blowup). Consider a unimodular integral $n \times n$ matrix $\sigma = (\sigma_{i,j})$, and define the *toric blowup* (or *modification*) associated toσ as the birational morphism

$$\begin{aligned} \pi_\sigma \colon (\mathbb{C}^*)^n &\to (\mathbb{C}^*)^n \\ (x_1, \dots, x_n) &\mapsto (x_1^{\sigma_{1,1}} x_2^{\sigma_{1,2}} \cdots x_n^{\sigma_{1,n}}, \dots, x_1^{\sigma_{n,1}} x_2^{\sigma_{n,2}} \cdots x_n^{\sigma_{n,n}}). \end{aligned}$$

In particular, if we have a regular simplicial cone in Σ^* of maximum dimension given by vectors $\{r_1,\dots, r_n\}$, we can consider the unimodular matrix $\sigma = (r_1 \quad r_2 \quad \cdots \quad r_n)$ and associate to it the birational map $\pi_\sigma \colon \mathbb{C}^n_\sigma \to \mathbb{C}^n$, which we ought to think about as one of the different charts of the resolution. With that, we construct a non-singular variety X as the quotient of the disjoint union $\bigsqcup_\sigma \mathbb{C}_\sigma$ over all the regular cones with the following identification. Two points $x \in \mathbb{C}^n_\sigma$ and $y \in \mathbb{C}^n_\tau$ are identified if, and only if, the birational map $\pi_{\tau^{-1}\sigma}$ is defined at the point x and $\pi_{\tau^{-1}\sigma}(x) = y$.

It can be verified that X is non-singular, and the maps $\{\pi_\sigma \colon \mathbb{C}^n_\sigma \to \mathbb{C}^n \mid \sigma \text{ regular simplicial cone}\}$ glue into a proper analytic map $\pi \colon X \to \mathbb{C}^n$.

Definition 3.10 (Associated toric blowup). The map $\pi \colon X \to \mathbb{C}^n$ is called the *toric blowup* (or modification) *associated* with Σ^* at the origin, where Σ^* is a regular simplicial cone subdivision of Σ.

Finally, we arrive at the result justifying our claim that the geometric properties of the Newton polyhedron contains the information of a resolution for NND singularities.

Theorem 3.11 ([21, p. 101]). *If f is Newton non-degenerate, then the associated toric blowup $\pi \colon X \to \mathbb{C}^n$ is a good resolution of f as a germ at the origin.*

In the same combinatorial spirit, it is possible to obtain a more explicit expression for the topological zeta function. For that, let us first introduce the following terms.

Definition 3.12. Let τ be a face in $\Gamma(f)$, and consider a decomposition of the associated cone $\Delta_\tau = \bigcup_{i=1}^r \Delta_i$ in simplicial cones of dimension $\dim \Delta_\tau = l$ such that $\dim(\Delta_i \cap \Delta_j) < l$, for all $i \neq j$. Then, define

$$J(\tau, s) := \sum_{i=1}^{r} J_{\Delta_i}(s), \quad \text{with} \quad J_{\Delta_i}(s) = \frac{\text{mult}(\Delta_i)}{(N(a_{i_1})s + k(a_{i_1})) \cdots (N(a_{i_l})s + k(a_{i_l}))}$$

being $a_{i_1}, \ldots, a_{i_l} \in \mathbb{N}^n$ the linearly independent primitive integral vectors that generate Δ_i. Lastly, if $\tau = \Gamma(f)$, we rather take $J(\tau, s) = 1$.

Remark 3.13. By [8, Lemme 5.1.1], the definition of $J(\tau, s)$ is independent of the choice of the decomposition of Δ_τ in simplicial cones.

Theorem 3.14 ([8, Théorème 5.3])**.** *Let $f \in \mathbb{C}[x]$ be a polynomial Newton non-degenerate, then*

$$Z(s) = \sum_{\substack{\tau \text{ vertex} \\ \text{of } \Gamma(f)}} J(\tau, s) + \left(\frac{s}{s+1}\right) \sum_{\substack{\tau \text{ face of } \Gamma(f) \\ \dim \tau \geq 1}} (-1)^{\dim \tau} (\dim \tau)! \, \text{Vol}(\tau) J(\tau, s).$$

As the poles of $Z(s)$ arise from the poles of the terms $J(\tau, s)$ in the sum, we see that they are still either -1 or of the form $-k(a)/N(a)$ (see also [16, Théorème 5.3.1]), which justifies the choice of notation.

3.3. Sketch of the proof and additional hypothesis

We are now ready to sketch the approach by Loeser in the proof of the monodromy conjecture for the Newton non-degenerate case. One relevant subtlety is, as stated in Remark 3.8, that the regular subdivision is not unique, and therefore so are the residue numbers. For that reason, Loeser decides to work with *toric* residue numbers computed directly from the original dual fan, without performing a regular subdivision.

Definition 3.15 (Toric residue numbers)**.** If τ, τ' are two distinct faces of codimension 1 of the Newton polyhedron at the origin of f, we denote by $\beta(\tau, \tau')$ the greatest common divisor of the minors of order 2 of the matrix $(a(\tau), a(\tau'))$, where $a(\tau)$ is a primitive integral vector defining the face τ. Additionally, set

$$\lambda(\tau, \tau') = k(\tau') - \frac{k(\tau)}{N(\tau)} N(\tau'), \quad \varepsilon(\tau, \tau') = \lambda(\tau, \tau')/\beta(\tau, \tau'),$$

whenever $N(\tau) \neq 0$, which is the case if τ is a compact face.

However, the drawback is that the resolution obtained from the original dual fan need not be minimal, and the subsequent computations require special care, leading to the introduction of this β factor which appears as the degree of a finite morphism between singular toric varieties.

Then, the positive result proven by Loeser is the following.

Theorem 3.16 ([16, Théorème 5.5.1])**.** *Let f be a comfortable polynomial verifying $f(0) = 0$, with Newton diagram $\Gamma(f)$, and Newton non-degenerate. Suppose that all compact faces τ_0 verify*

(i) $\frac{k(\tau_0)}{N(\tau_0)} < 1$.

(ii) For every face τ of codimension 1 of $\Gamma(f)$, distinct of τ_0 and having non-empty intersection with τ_0, we have $\varepsilon(\tau_0, \tau) \notin \mathbb{Z}$.

Then, the real parts of the poles of the zeta function of f are roots of the Bernstein–Sato polynomial of f.

As explained, this is based on a cohomological argument of existence of a non-zero class, which in turn forces the two stated conditions, the second one basically ensuring that the monodromies are not identity.

Remark 3.17*.* Loeser already points out in [16, Remarque 5.5.2.1] that if one replaces the condition $\frac{k(\tau_0)}{N(\tau_0)} < 1$ with $\frac{k(\tau_0)}{N(\tau_0)} \notin \mathbb{N}$, this is enough to prove the weak version of the conjecture.

As for the second hypothesis, although it is not clear if it is possible to remove, we can try to relax it. Indeed, one ought to expect that non-positive residue numbers could be allowed to happen, after possibly generalizing results in the spirit of [7, Proposition 2.14] or [4, Proposition 11.1]. Therefore, we will next try to construct examples where this is the case, and analyze the contributions of such divisors violating the condition.

3.4. Constructing counterexamples to the second condition

To begin, it should be mentioned that all examples studied have been checked to satisfy the topological strong monodromy conjecture (and thus the weak version too). Nonetheless, the relevant findings are examples where the second condition on the (toric) residue numbers is not met, and even more, where ε is a positive integer.

Even more interestingly, these examples have been motivated geometrically from the Newton polyhedron. More precisely, f is first constructed by introducing monomials $x^p + y^q + z^r$ for p, q, r large enough integers. Then, adding small mixed monomials of the type $x^s y^t z^u$ for small enough integers s, t, u, we obtain small compact faces close to the origin. By choosing the exponents appropriately, we can construct a polyhedron whose faces have normal vectors as desired. In particular, we can find pairs of adjacent faces for which the corresponding divisors of the resolution give rise to (toric) residue numbers that are integers.

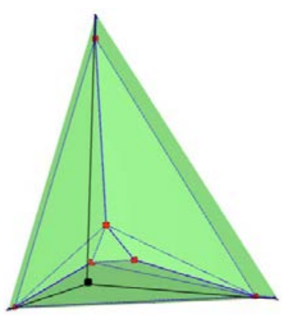

Figure 2: Example of the construction of $\Gamma(x^5 + y^6 + z^7 + x^2yz + xyz^2 + xy^2z)$.

As an illustrative example, we consider $f = x^5 + y^6 + z^7 + x^2yz + xyz^2 + xy^2z$, whose Newton diagram is depicted in the above Figure 2. The original rays in the dual fan are:

$$[(0,0,1),(0,1,0),(1,0,0),(1,1,1),(1,1,2),(1,2,1),(3,1,1),(6,5,14),(7,18,5),(23,7,6)],$$

and a regular subdivision requires almost 400 new rays. Also, the Bernstein–Sato polynomial is

$$b_f(s) = \left(s+\tfrac{11}{6}\right)\left(s+\tfrac{9}{5}\right)\left(s+\tfrac{12}{7}\right)\left(s+\tfrac{5}{3}\right)\left(s+\tfrac{8}{5}\right)\left(s+\tfrac{11}{7}\right)\left(s+\tfrac{3}{2}\right)\left(s+\tfrac{10}{7}\right)\left(s+\tfrac{7}{5}\right)$$
$$\cdot\left(s+\tfrac{4}{3}\right)\left(s+\tfrac{9}{7}\right)\left(s+\tfrac{5}{4}\right)\left(s+\tfrac{6}{5}\right)\left(s+\tfrac{7}{6}\right)\left(s+\tfrac{8}{7}\right)(s+1)^3\left(s+\tfrac{6}{7}\right)\left(s+\tfrac{5}{6}\right)\left(s+\tfrac{4}{5}\right)\left(s+\tfrac{3}{4}\right),$$

and the (local) topological zeta function (which is the same expression as the global version Theorem 3.14, but the second sum runs only over compact faces) is

$$Z_0(s) = \frac{81/4}{s+3/4} - \frac{72/5}{s+4/5} - \frac{70/6}{s+5/6} - \frac{48/7}{s+6/7} + \frac{14}{s+1},$$

with poles $\{-3/4, -4/5, -5/6, -6/7, -1\}$. Therefore, it is immediate that the strong monodromy conjecture holds for this case.

However, we have the ray $(1,1,2)$ with candidate pole value $\sigma = -k/N = -4/5$ appearing, and which is a bad divisor. Indeed, we have the (toric) residue number

$$\varepsilon((1,1,2),(7,8,5)) = 2.$$

In light of this, the next natural question is whether we can discard the contribution of this *bad* divisor to the zeta function, and work only with those satisfying both conditions (in the spirit of the case of plane curves, where a study of the dual graph allowed to discard non-rupture divisors for which we don't have the cohomological result).

Following the above example, the two divisors with candidate pole value $-4/5$ are the bad divisor $(1,1,2)$ and also $(1,2,1)$. So we compute the contributions of each divisor to the (local) topological zeta function, by summing only the terms from cones Δ_i that contain the ray as one of its generating rays. As a warning, for a ray a this is not the same as simply taking the terms where a fraction $\frac{1}{N(a)s+k(a)}$ appears, as it can happen that another ray a' gives rise to the same candidate pole.

In particular, we find that the residues of the individual contributions are

$$\operatorname*{Res}_{s=-4/5} Z_{0;(1,2,1)}(s) = -\frac{47}{5}, \quad \operatorname*{Res}_{s=-4/5} Z_{0;(1,1,2)}(s) = -\frac{47}{5}$$

so we can't discard the bad divisor.

Remark 3.18. Another remark to point here is that the total residue won't necessarily be the sum of these residues. Indeed, there can be cones where both rays appear as generating rays (this happens precisely if the associated divisors intersect), and in that case we would need to subtract the doubly counted contribution. More generally, an inclusion-exclusion formula should be applied in order to compare the separate contributions and the total one for divisors with the same candidate pole value.

Altogether, the reasoning and study of the presented example confirms that we can't omit the additional hypothesis on the residue numbers required for the proof of the conjecture in the Newton non-degenerate case and, even worse, that the possible divisors not satisfying it can indeed contribute with non-zero residue to the poles of the topological zeta function. In other words, the approach by Loeser based on the construction of a non-zero cohomology class via the mentioned results does not allow to extend the proof of the conjecture to the general case.

Acknowledgements

The author was partially supported by CFIS Mobility Program, an Erasmus+ grant co-funded by the European Union and a Santander Erasmus scholarship. The author would like to thank Guillem Blanco Fernández and Josep Àlvarez Montaner for their supervision of the project, and KU Leuven for hosting him during its development.

References

[1] V.I. Arnold, S.M. Gusein-Zade, A.N. Varchenko, *Singularities of Differentiable Maps. Volume 2. Monodromy and Asymptotics of Integrals*, Translated from the Russian by Hugh Porteous and revised by the authors and James Montaldi, Reprint of the 1988 translation, Mod. Birkhäuser Class. Birkhäuser/Springer, New York, 2012.

[2] J.H. Bernstein, Analytic continuation of generalized functions with respect to a parameter, *Functional Anal. Appl.* **6(4)** (1972), 273–285.

[3] J.-E. Björk, Dimensions over algebras of differential operators, Département de mathématiques (1973).

[4] G. Blanco, Bernstein–Sato polynomial of plane curves and Yano's conjecture, PhD Thesis, Universitat Politècnica de Catalunya, 2020.

[5] F.J. Castro Jiménez, Modules over the Weyl algebra, in: *Algebraic Approach to Differential Equations*, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2010, pp. 52–118.

[6] D.A. Cox, J.B. Little, H.K. Schenck, *Toric Varieties*, Grad. Stud. Math. **124**, American Mathematical Society, Providence, RI, 2011.

[7] P. Deligne, G.D. Mostow, Monodromy of hypergeometric functions and nonlattice integral monodromy, *Inst. Hautes Études Sci. Publ. Math.* **63** (1986), 5–89.

[8] J. Denef, F. Loeser, Caractéristiques d'Euler–Poincaré, fonctions zêta locales et modifications analytiques, *J. Amer. Math. Soc.* **5(4)** (1992), 705–720.

[9] H. Esnault, E. Viehweg, *Lectures on Vanishing Theorems*, DMV Sem. **20**, Birkhäuser Verlag, Basel, 1992.

[10] W. Fulton, *Introduction to Toric Varieties*, Ann. of Math. Stud. **131**, William Roever Lectures Geom., Princeton University Press, Princeton, NJ, 1993.

[11] I. Gelfand, Some aspects of functional analysis and algebra, in: *Proceedings of the International Congress of Mathematicians*, Vol. 1 (Amsterdam, 1954), Erven P. Noordhoff N. V., Groningen, 1957, pp. 253–276.

[12] H. Hironaka, Resolution of singularities of an algebraic variety over a field of characteristic zero: II, *Ann. of Math. (2)* **79(2)** (1964), 205–326.

[13] M. Kashiwara, *B*-functions and holonomic systems. Rationality of roots of *B*-functions, *Invent. Math.* **38(1)** (1976/77), 33–53.

[14] G. Kempf, F. Knudsen, D. Mumford, B. Saint-Donat, *Toroidal Embeddings 1*, Lecture Notes in Math. **339**, Springer-Verlag, Berlin-New York, 1973.

[15] F. Loeser, Fonctions d'Igusa *p*-adiques et polynômes de Bernstein, *Amer. J. Math.* **110(1)** (1988), 1–21.

[16] F. Loeser, Fonctions d'Igusa *p*-adiques, polynômes de Bernstein, et polyèdres de Newton, *J. Reine Angew. Math.* **412** (1990), 75–96.

[17] J. MacLaurin, The resolution of toric singularities, PhD Thesis, School of Mathematics, The University of New South Wales, 2006.

[18] B. Malgrange, Sur les polynômes de I. N. Bernstein, in: *Séminaire Goulaouic–Schwartz 1973–1974: Équations aux dérivées partielles et analyse fonctionnelle*, Exp. No. 20, École Polytechnique, Centre de Mathématiques, Paris, 1974, 10 pp.

[19] B. Malgrange, Intégrales asymptotiques et monodromie, *Ann. Sci. École Norm. Sup. (4)* **7** (1974), 405–430.

[20] B. Malgrange, Polynômes de Bernstein–Sato et cohomologie évanescente, *Astérisque* **101-102** (1983), 243–267.

[21] M. Oka, Geometry of plane curves via toroidal resolution, in: *Algebraic Geometry and Singularities* (La Rábida, 1991), Progr. Math. **134**, Birkhäuser Verlag, Basel, 1996, pp. 95–121.

[22] W. Veys, Introduction to the monodromy conjecture, in: *Handbook of Geometry and Topology of Singularities VII*, Springer, Cham, 2025, pp. 721–765.

REPORTS@SCM

AN ELECTRONIC JOURNAL OF THE
SOCIETAT CATALANA DE MATEMÀTIQUES

Existence of non-convex rotating vortex patches to the 2D Euler equation

Gerard Castro López
ETH Zurich
gcastro@student.ethz.ch

Resum *(CAT)*
Els únics exemples explícits de vòrtexs en rotació uniforme a les equacions d'Euler 2D són els cercles i les el·lipses. Els altres exemples dels quals es coneixen propietats quantitatives són propers a aquests.
En aquest article presentem l'existència de vòrtexs no convexos, lluny dels règims pertorbatius, podent obtenir-ne una descripció quantitativa precisa. Per demostrar-ho utilitzem una combinació d'anàlisi i tècniques de demostració assistida per ordinador.

Abstract *(ENG)*
The only explicit examples of uniformly rotating vortex patches to the 2D Euler equations are circles and ellipses. The other examples for which quantitative properties are known are close to these ones.
In this paper we present the existence of non-convex ones, far from the perturbative regimes, being able to obtain a precise quantitative description. To prove it we use a combination of analysis and computer assisted proofs techniques.

Societat Catalana de Matemàtiques

Institut d'Estudis Catalans

Keywords: *PDE, fluid mechanics, 2D Euler, vortex patches, computer assisted proofs.*
MSC (2020): *Primary 35Q31, 57P10. Secondary 65G30, 76B03.*

Received: *July 2, 2025.*
Accepted: *August 10, 2025.*

1. Introduction

The behavior of ideal, incompressible fluids is governed by the Euler equations, a set of partial differential equations derived by Leonhard Euler in the 1750s. In two dimensions, these equations can be formulated in terms of vorticity, ω, which measures the local rotation of the fluid in the following way:

$$\begin{cases} \partial_t \omega + (K * \omega) \cdot \nabla \omega = 0 & \text{in } \mathbb{R}^2 \times [0, \infty), \\ \omega(\cdot, 0) = \omega_0(\cdot) & \text{in } \mathbb{R}^2, \end{cases}$$

where $K(\mathbf{y}) = \frac{\mathbf{y}^\perp}{2\pi|\mathbf{y}|^2}$. A special class of solutions, known as *vortex patches*, occurs when the initial vorticity is constant within a defined region D_0 and zero elsewhere. The vorticity remains constant along particle trajectories, so the patch simply deforms over time. A more exhaustive introduction with more detailed computations can be found in [10].

This study focuses on a particular type of vortex patch called a *V-state*, which is a patch that rotates with a uniform angular velocity (Ω) without changing its shape. Mathematically, if a patch is defined by a domain D_0 at time $t = 0$, its evolution is given by $D(t) = M(\Omega t)D_0$, where M is a rotation matrix.

The boundary of the patch ∂D_0, parametrized in polar coordinates by the function $R(\alpha)$, must satisfy the following non-local integro-differential equation:

$$R(\alpha)R'(\alpha) = F[R], \tag{1}$$

where $F[R]$ is an integral operator given by:

$$F[R] := \frac{1}{4\pi\Omega} \int_0^{2\pi} \cos(\alpha-\beta) \log\left((R(\alpha)-R(\beta))^2 + 4R(\alpha)R(\beta)\sin^2\left(\frac{\alpha-\beta}{2}\right)\right)(R(\alpha)R'(\beta) - R'(\alpha)R(\beta))\, d\beta$$
$$+ \frac{1}{4\pi\Omega} \int_0^{2\pi} \sin(\alpha-\beta) \log\left((R(\alpha)-R(\beta))^2 + 4R(\alpha)R(\beta)\sin^2\left(\frac{\alpha-\beta}{2}\right)\right)(R(\alpha)R(\beta) + R'(\alpha)R'(\beta))\, d\beta.$$

1.1. Previous work

The circle is a trivial solution, and in 1874, Kirchhoff ([9]) proved that ellipses are also V-states. Apart from these, no other explicit solutions are known. In the 1970s and 80s, numerical studies by Deem–Zabusky [3] and others revealed families of V-states with m-fold symmetry bifurcating from the circle, suggesting a rich variety of solutions beyond the classical examples.

Theoretical progress followed, with Burbea [1] and Hmidi, Mateu, and Verdera [6] proving the existence of local bifurcation branches from the disk for every integer symmetry $m \geq 3$. More recently, global bifurcation curves were constructed by Hassainia, Masmoudi, and Wheeler [5]. However, these powerful theoretical tools provide existence but lack quantitative information about the solutions far from the initial bifurcation point. This is a critical limitation because proving properties like non-convexity requires precise knowledge of the solution's shape. A general argument for all m-fold branches will fail, as the branch for $m = 2$ is known to be convex and for $m = 3$ it is expected.

This paper addresses this gap by providing the first positive existence result with quantitative information for a V-state far from the perturbative regions around the circle or ellipses. We prove the existence of a non-convex, 6-fold symmetric V-state.

Due to length constraints, most proofs are omitted from this article. A more thorough explanation of the results is detailed by Gómez-Serrano and the present author in [2].

1.2. Results

Theorem 1.1 (Main theorem)**.** *There exists an analytic solution $R(x)$ of the V-state equation (1), with $\Omega = \frac{1537}{3750}$, such that its associated vortex patch $D \subset \mathbb{R}^2$ is non-convex and has 6-fold symmetry. See Figure 1.*

Corollary 1.2. *There exists $\delta > 0$ such that for any angular velocity in $(\Omega - \delta, \Omega + \delta)$ there exists an analytic solution R to (1) with 6-fold symmetry, where Ω is the angular velocity given by Theorem 1.1.*

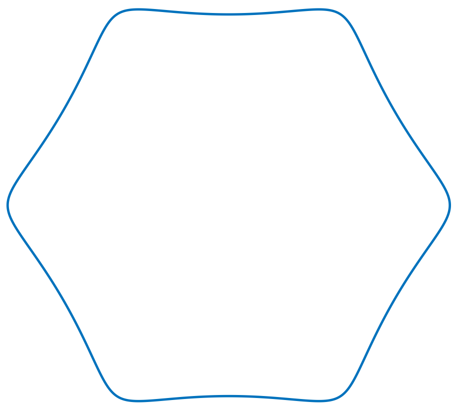

Figure 1: The boundary ∂D is a curve contained in the plotted line.

Proof of Theorem 1.1. In Section 2 we prove the existence of a solution close to R_0, in Section 3 we prove the analyticity and non-convexity and in the last Section 4 we explain the computer assisted parts of the proof. □

2. Existence

Our approach is to formulate the problem as a fixed point equation and solve it using a combination of analytical estimates and rigorous numerical computations. The core idea is to find an approximate solution and then prove that a true solution exists in its vicinity.

2.1. Fixed point equation around the approximate solution

The first step is to find a highly accurate approximate solution, $R_0(x)$. We define it as a truncated Fourier series with 6-fold symmetry ($m = 6$):

$$R_0(x) := \sum_{k=0}^{N_0} c_k \cos(kmx)$$

with $N_0 = 30$. The coefficients c_k and the angular velocity Ω are chosen to make the error of this approximation, $E[0](x) = R_0'(x)R_0(x) - F[R_0](x)$, as small as possible. We seek a true solution $R(x)$ as a perturbation of $R_0(x)$:

$$R(x) = R_0(x) + v(x) \quad \text{where} \quad v(x) = \int_0^x \tilde{u}(y)\, dy. \tag{2}$$

Here, u is a function in $L^2([0, \pi/m])$ and $\tilde{u}$ is its odd, $2\pi/m$ periodic extension. This formulation is designed to fix the scaling and rotation symmetries of the problem while preserving the m-fold symmetry.

Taking the Fréchet derivative of the equation (1), we can write the equation for the perturbation as $\mathcal{L}v = E[0] + \mathcal{N}_L[v]$, where $\mathcal{L}$ is a linear operator and $\mathcal{N}_L[v] = O(v^2)$ is nonlinear. Using the expression (2) and the symmetries of $\tilde{u}$, we write this equation in terms of u in the following way:

$$Lu(x) = E[0](x) + N_L[u](x) \quad \forall x \in [0, \pi/m].$$

- L is a linear operator defined as $Lu := u + \int_0^{\pi/m} K(x, y)u(y)\, dy$, where K is a $L^2\left([0, \frac{\pi}{m}]^2\right)$ function that depends on the approximate solution R_0 in a nontrivial way.
- E_0 is the error of the approximate solution R_0 as we have mentioned before.
- N_L is a nonlinear operator that contains higher-order terms of the perturbation u.

If we prove that L is invertible, the problem is now reduced to showing that the operator $Gu := L^{-1}(E[0] + N_L[u])$ has a fixed point in a small ball around the origin in an appropriate function space. We choose the space $L^2([0, \pi/m])$ and seek a solution u within a ball of radius $\epsilon = 3 \cdot 10^{-5}$.

2.2. Inverting the linear operator

A direct inversion of the operator L is not feasible, since it depends very nonlinearly (and non-locally) on R_0. Instead, we use a computer assisted approach. The main idea to prove the invertibility of L is to first approximate L by $L_F = \text{Identity} + \text{Finite Rank}$, then prove the invertibility of L_F and finally prove that the approximation error $L - L_F$ is small enough, making L invertible via a Neumann series.

Definition 2.1. Let $\{e_n(x)\}_n$ be the normalized Fourier basis of $X_m = L^2\left([0, \frac{\pi}{m}]\right)$. Also let $N = 201$, and $E_N = \text{span}\{e_n\}_{n=1}^N$ be the subspace generated by the first N vectors. Similarly, let $E_N^\perp$ be its orthogonal subspace. We will also define $L_F = I + \mathcal{K}_F \colon X_m \mapsto X_m$ where $\mathcal{K}_F \colon E_N \to E_N$ is given by

$$\mathcal{K}_F[u] = \int_0^{\frac{\pi}{m}} K_F(x, y)u(y)\, dy \quad \text{with} \quad K_F(x, y) := \sum_{k,l=1}^N A_{k,l}e_k(x)e_l(y),$$

where $A_{k,l}$ is finite explicit matrix very close to the projection of the operator $K(x, y)$ to the E_N subspace.

The proof of the next lemma is computer assisted and will be explained in Section 4.

Lemma 2.2. *The matrix $I + A$ is invertible and satisfies $\|(I + A)^{-1}\|_2 \le C_2$, with $C_2 := 8.8$.*

Lemma 2.3. *The operator L_F is invertible and $\|L_F^{-1}\|_2 < C_2$ with $C_2 := 8.8$.*

Proof. Let P_N be the projection operator onto E_N. Using that $\mathcal{K}_F = P_N\mathcal{K}_F = \mathcal{K}_F P_N$, L_F decouples in $E_N \bigoplus E_N^\perp$ as

$$L_F = \begin{pmatrix} I_{E_N} + A & \mathbf{0} \\ \mathbf{0} & I_{E_N}^\perp \end{pmatrix};$$

then as E_N is a finite dimensional vector space, to invert $I_{E_N} + A$ we have to invert the corresponding matrix, and the identity in $E_N^\perp$ is trivially inverted, so

$$L_F^{-1} f := (I_{E_N} + A)^{-1} P_N f + (I - P_N) f.$$

We can conclude that L_F is invertible. Moreover

$$\begin{aligned} \|L_F^{-1} f\|_{L^2} = \|P_N L_F^{-1} f\|_{L^2} + \|(I - P_N) L_F^{-1} f\|_{L^2} &\leq \|(I_{E_N} + A)^{-1}\|_2 \|P_N f\|_{L^2} + \|(I - P_N) f\|_{L^2} \\ &\leq \max\{\|(I_{E_N} + A)^{-1}\|_2, 1\} \|f\|_{L^2}, \end{aligned}$$

hence its norm is bounded by C_2 because by Lemma 2.2, $\|(I + A)^{-1}\|_2 \leq C_2$ and $1 < C_2$. □

The proof of the next lemma is again computer assisted and it will be explained in Section 4.

Lemma 2.4. *The error of approximating the operator L by L_F satisfies $\|L - L_F\|_2 \leq C_3$, with $C_3 := 0.085$.*

We can now state and prove the main result of this subsection.

Proposition 2.5. *The linear operator $L\colon X_m \to X_m$ is invertible. Moreover $\|L^{-1}\|_2 \leq C_1 = 35$.*

Proof. Using that L_F is invertible, we can write

$$L = L_F(I + L_F^{-1}(L - L_F)).$$

We can then invert $I + L_F^{-1}(L - L_F)$ using a Neumann series because due to Lemmas 2.3, 2.4 we have that $\|L_F^{-1}(L - L_F)\|_2 \leq C_2 C_3 < 1$. As L_F is also invertible, we can conclude that L is invertible and

$$\|L^{-1}\|_2 = \|(I + L_F^{-1}(L - L_F))^{-1} L_F^{-1}\|_2 \leq \frac{C_2}{1 - C_2 C_3} = \frac{8.8}{0.252} < 35 = C_1.$$ □

2.3. Solving the fixed point equation

The goal is to use Banach Fixed Point theorem to prove the existence of solutions. For this we need control over the Lipschitz norm of G. We are going to state a proposition that together with the estimates on L^{-1} is going to allow us to prove the existence of a fixed point.

Proposition 2.6. *We have the following bounds on the nonlinear terms:*

(i) $\|E[0]\|_{L^2} \leq \epsilon_0$,

(ii) $\|N_L[u]\|_{L^2} \leq C_5 \|u\|_{L^2}^2$,

(iii) $\|N_L[u_1] - N_L[u_2]\|_{L^2} \leq \epsilon C_6 \|u_1 - u_2\|_{L^2}$

for $\epsilon_0 := 3 \cdot 10^{-7}$ and some explicit positive constants C_5, C_6.

The proof of the bound on the error of the approximate solution, the first bound of last lemma, is computer assisted. The rest of them are estimated by hand.

The next proposition is the main one of this section.

Proposition 2.7. *The operator $G[u] := L^{-1}(E[0] + N_L[u])$ has a fixed point in the ball of radius ϵ.*

Proof. We check that the operator maps the ball $B_\epsilon(0)$ into itself:

$$\|G[u]\|_{L^2} \leq C_1(\epsilon_0 + C_5\epsilon^2) < \epsilon,$$

and that it is Lipschitz

$$\|G[u_1] - G[u_2]\|_{L^2} \leq C_1 C_6 \epsilon \|u_1 - u_2\|_{L^2} < \|u_1 - u_2\|_{L^2}.$$

In both inequalities we have checked that the explicit values of the constants satisfy them. We conclude by Banach Fixed Point theorem. □

Once we have proven the existence of this solution it is possible to check that indeed $R = R_0 + \int_0^x \tilde{u}$ satisfies equation (1).

3. Non-convexity and improved regularity

3.1. Proof of non-convexity

To prove this, we first note that using Hölder inequality and the bound $\|u\|_{L^2} \leq \epsilon$ we get the following enclosure for the solution $R(x)$

$$R_{\inf} := R_0(x) - \sqrt{\frac{\pi}{m}|\sin(mx)|} \leq R(x) \leq R_0(x) - \sqrt{\frac{\pi}{m}|\sin(mx)|} =: R_{\sup}(x).$$

The quantitative control we have on the solution allows us to rigorously prove its non-convexity. Let's define P_0, P_1 the points on the boundary of the patch at angle 0, $2\pi/m$. We will then prove that $R(\pi/m) < \left|\frac{P_0+P_1}{2}\right|$ so the midpoint does not belong to the domain of the patch D.

$$R(\pi/m) \leq R_{\sup}(\pi/m) = R_0\left(\frac{\pi}{m}\right) + \epsilon\sqrt{\frac{\pi}{m}} < R_0(0)\cos\left(\frac{\pi}{m}\right) = \left|\frac{P_0 + P_1}{2}\right|,$$

where the explicit inequality in the middle is checked using interval arithmetic with the computer.

3.2. Further regularity

We initially proved that R is an H^1 function. However, we can show that it is much smoother.

1. C^∞ **Regularity**: We use a bootstrapping argument. By rewriting the V-state equation, we show that if the k-th derivative $\partial^k R$ is bounded, then $\partial^{k+1} R$ is also bounded, and by induction, R is C^∞. The argument involves carefully differentiating the non-local equation and controlling the singularities in the integral kernels to obtain the following inequality:

$$\|\partial^{k+2} R\|_{L^\infty} \lesssim 1 + \delta(1 + \|\partial^{k+1} R\|_{L^\infty}^{\beta_k})\|\partial^{k+2} R\|_{L^\infty} + \frac{1}{\delta}\|\partial^{k+1} R\|_{L^\infty}^{\gamma_k},$$

   for a sufficiently small $\delta > 0$.

2. **Analyticity**: To prove that the boundary is analytic, we rephrase the problem as a free boundary elliptic problem for the stream function ψ. Then we can explicitly check that the gradient is not vanishing in the boundary. Once we have this the analyticity is a direct consequence due to [8, Theorem 3.1'].

4. Computer assisted estimates

The proof of the main theorem has strongly relied in using the computer to perform difficult computations. We will explain in which steps and how we were able to obtain the required bounds. We refer to [4] for a thorough overview of computer assisted proofs in PDEs.

To perform rigorous computations with a computer, we cannot use floating-point arithmetic, as the results would not be rigorous because we have no control over the rounding error. Therefore, we must use interval arithmetic. Interval arithmetic treats numbers as intervals, where the interval bounds are numbers representable with a fixed number of digits. In this arithmetic, operations are defined between intervals in such a way that for all numbers belonging to an input interval, the result of the operation is guaranteed to be included in the final output interval. This method allows for the propagation of errors to ultimately obtain an interval that contains the exact result.

We used the `C++` library `Arb` for these computations [7]. The key computer assisted steps were:

- **Bounding the matrix inverse**: Finding a rigorous upper bound for $\|(I + A)^{-1}\|_2$ by computing a lower bound for the minimum eigenvalue of the symmetric matrix $(I + A)^T(I + A)$.

- **Bounding integrals**: Using high-order quadrature rules with rigorous error bounds to compute norms of functions involving difficult integrals, such as the error of the approximate solution $\|E[0]\|_{L^2}$ and the approximation error of the linear operator $\|K - K_F\|_{L^1_x L^\infty_y}, \|K - K_F\|_{L^1_y L^\infty_x}$. Special care was taken to handle the logarithmic singularities in the integrands.

- **Verifying final inequalities**: Checking the conditions for the Banach Fixed Point theorem and the final non-convexity inequality by plugging in the rigorously computed bounds for all constants.

The numerics required a nontrivial amount of computing size; for instance, bounding the error of the approximate solution to the order of 10^{-7} took about 9 hours in 64 CPUs, and the bounding of the error of the kernel approximation took 29 hours in also 64 CPUs.

Acknowledgements

I would like to express my sincere gratitude to Javier Gómez Serrano for his outstanding supervision of my Bachelor's thesis, which led to this article. I also wish to wholeheartedly thank the Department of Mathematics at Brown University for hosting me during the development of this project.

This article is published in connection with the Emmy Noether Prize. I am deeply honored to be a recipient and thank the jury for their work and consideration.

I acknowledge the CFIS Mobility Program for partially funding this research, with special thanks to the Fundació Privada Mir-Puig, the CFIS partners, and the donors of the crowdfunding program. I was partially supported by a MOBINT grant from the Generalitat de Catalunya, an Erasmus+ grant co-funded by the European Union, and NSF grants DMS-2245017, DMS-2247537, and DMS-2434314. I also thank Brown University for providing the computing facilities used in this research.

References

[1] J. Burbea, Motions of vortex patches, *Lett. Math. Phys.* **6(1)** (1982), 1–16.

[2] G. Castro-López, J. Gómez-Serrano, Existence of analytic non-convex V-states, *Comm. Math. Phys.* **406(9)** (2025), Paper no. 217.

[3] G.S. Deem, N.J. Zabusky, Vortex waves: stationary "V states," interactions, recurrence, and breaking, *Phys. Rev. Lett.* **40(13)** (1978), 859–862.

[4] J. Gómez-Serrano, Computer-assisted proofs in PDE: a survey, *SeMA J.* **76(3)** (2019), 459–484.

[5] Z. Hassainia, N. Masmoudi, M.H. Wheeler, Global bifurcation of rotating vortex patches, *Comm. Pure Appl. Math.* **73(9)** (2020), 1933–1980.

[6] T. Hmidi, J. Mateu, J. Verdera, Boundary regularity of rotating vortex patches, *Arch. Ration. Mech. Anal.* **209(1)** (2013), 171–208.

[7] F. Johansson, Arb: efficient arbitrary-precision midpoint-radius interval arithmetic, *IEEE Trans. Comput.* **66(8)** (2017), 1281–1292.

[8] D. Kinderlehrer, L. Nirenberg, Regularity in free boundary problems, *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)* **4(2)** (1977), 373–391.

[9] G. Kirchhoff. *Mechanik*, Vorlesungen über mathematische Physik, Vol. 1, B. G. Teubner, Leipzig, 1874.

[10] A.J. Majda, A.L. Bertozzi, *Vorticity and Incompressible Flow*, Cambridge Texts Appl. Math. **27**, Cambridge University Press, Cambridge, 2002.

REPORTS@SCM

AN ELECTRONIC JOURNAL OF THE
SOCIETAT CATALANA DE MATEMÀTIQUES

Ideals of p^e-th roots of plane curves in positive characteristic

Pedro López Sancha
Universitat Politècnica de Catalunya
pedro.lopez.sancha@upc.edu

Resum *(CAT)*
Una estratègia per estudiar varietats algebraiques és construir invariants algebraics que mesurin les seves singularitats. Sobre els nombres complexos, destaquen els ideals multiplicadors i els nombres de salt. En característica positiva, les seves contraparts són els ideals de test i els F-nombres de salt. En aquest projecte, calculem els ideals de test i F-nombres de salt de corbes planes quasi-homogènies, així com de les seves deformacions a nombre de Milnor constant, per una quantitat infinita de característiques $p > 0$. En aquests casos, veiem que els ideals de test són la reducció mòdul p dels ideals multiplicadors.

Abstract *(ENG)*
A common approach to studying algebraic varieties is through algebraic invariants that measure their singularities. Over the complex numbers, a celebrated example of such invariants include the multiplier ideals and the jumping numbers. In positive characteristic, their counterparts are the test ideals and F-jumping numbers. In this work, we compute the test ideals and F-jumping numbers of quasi-homogeneous plane curves, as well as their one-monomial constant Milnor number deformations, for infinitely many characteristics $p > 0$. In these cases, we see that the test ideals are the modulo p reduction of the multiplier ideals.

Keywords: *test ideals, F-jumping numbers, quasi-homogeneous plane curve, constant Milnor number deformations.*
MSC (2020): *Primary 13A35, 14F10. Secondary 14B05.*

Received: *July 2, 2025.*
Accepted: *July 18, 2025.*

Societat Catalana de Matemàtiques

Institut d'Estudis Catalans

1. Introduction

A challenge that lies at the heart of modern algebraic geometry is the classification of algebraic varieties which, in particular, encompasses the characterization of their singularities. The most common approach to this problem is to construct algebraic and geometric invariants to quantify the singularities.

In birational algebraic geometry over the complex numbers, or more generally, over fields of characteristic zero, one can take advantage of the existence of a resolution of singularities to construct invariants. A celebrated example of such invariants is the family of multiplier ideals. Given a hypersurface defined by the vanishing locus of a polynomial f, the multiplier ideals $\mathcal{J}(f^\lambda)$ form a family of ideals indexed by nonnegative real numbers $\lambda \in \mathbb{R}_{\geq 0}$. These give a descending chain of ideals $\mathcal{J}(f^\lambda) \supseteq \mathcal{J}(f^\mu)$ whenever $\lambda \leq \mu$, which in addition is right-semicontinuous, meaning that $\mathcal{J}(f^\lambda) = \mathcal{J}(f^{\lambda+\varepsilon})$ for some $\varepsilon > 0$. The values $\lambda > 0$ where the chain jumps, that is, $\mathcal{J}(f^{\lambda-\varepsilon}) \supsetneq \mathcal{J}(f^\lambda)$ for arbitrarily small $\varepsilon > 0$, are known as the jumping numbers of f. The smallest jumping number among them is called the log-canonical threshold. By means of the resolution of singularities, one can show this set rational and discrete. References are made to [6].

The multiplier ideals and jumping numbers encode the singularities of the hypersurface determined by f in subtle ways. Suppose, for instance, that the vanishing locus of f is a curve C in the complex plane with a singularity at the origin. If one deforms C into a new curve C' while preserving the analytic type of the singularity, then the entire family of multiplier ideals remains unchanged. On the contrary, if the deformation only preserves the topological type, then the jumping numbers still agree, although the multiplier ideals may, in general, differ.

Over fields of positive characteristic $p > 0$, there is no resolution of singularities available for varieties of arbitrary finite dimension. In this setting, the Frobenius endomorphism, or p-th power map, serves as a substitute tool. The test ideals, which play the role of the multiplier ideals, were introduced by Hochster and Huneke as an auxiliary tool in tight closure theory [5], and were later refined by Hara and Yoshida [4]. We shall adopt the construction of Blickle, Mustață, and Smith (see Definition 2.9), which generalizes earlier definitions [3]. In brief, the test ideals $\tau(f^\lambda)$ of a polynomial f are a nested, right-semicontinuous family of ideals indexed over the nonnegative real numbers $\lambda \in \mathbb{R}_{\geq 0}$. The spots where the chain of test ideals "jumps" are the F-jumping numbers of f, the smallest of which is the F-pure threshold.

It is a well-established fact due to Mustață, Takagi, and Watanabe, that if f is a polynomial defined over the integers, the log-canonical threshold of f can be recorvered from the F-pure thresholds of the reductions f_p of f modulo a prime p, as $p \to \infty$ [9]. A profound conjecture in arithmetic geometry, the weak ordinarity conjecture, proposes a further connection, namely, that the test ideals of the reductions f_p can be calculated by reducing the multiplier ideals modulo p, for all primes in a Zariski-dense set [8]. In this sense, the theories of multiplier ideals in characteristic zero and test ideals in positive characteristic are closely analogous.

Test ideals and F-jumping numbers are notoriously difficult to compute. In the few cases where explicit descriptions are known—such as elliptic curves, diagonal hypersurfaces, determinantal ideals of maximal minors, or ideals invariant under the action of a subgroup of a linear group, the calculations rely on the arithmetic or combinatorial properties of the variety. A naive yet effective approach to obtain $\tau(f^\lambda)$ is to calculate the ideals p^e-th roots of f, denoted $\mathcal{C}_R^e \cdot f^n$ (see Definition 2.5), for a fixed integer $e \geq 0$. As $n \geq 0$ ranges over the natural numbers integers, one obtains descending chain of ideals

$$\mathcal{C}_R^e \cdot f^0 \supseteq \mathcal{C}_R^e \cdot f \supseteq \mathcal{C}_R^e \cdot f^2 \supseteq \cdots \supseteq \mathcal{C}_R^e \cdot f^n \supseteq \mathcal{C}_R^e \cdot f^{n+1} \supseteq \cdots$$

which, in essence, contains all the test ideals of f, and codifies the F-jumping numbers.

In this work, we begin by studying quasi-homogeneous plane curves C over a perfect field K of characteristic $p > 0$, that is, curves in K^2 given as the vanishing locus of a polynomial of the form $f = x^a + y^b$, with $a, b \geq 2$. For these, we observe the ideals of p^e-th roots, and consequently the test ideals are monomial ideals for sufficiently big characteristics $p \gg 0$. To determine the F-jumping numbers, we pose a linear integer programming problem, and provide its solution.

We then turn to deformations C' of the original curve C, which are curves given as the zeros of polynomials $g = f + \sum_i t_i x^{\alpha_i} y^{\beta_i}$. We restrict ourselves, however, to one-monomial deformations $g = f + tx^\alpha y^\beta$. From the algebro-geometric standpoint, it is most natural to consider deformations that preserve the singularity type of C at the origin, namely, constant Milnor number deformations. In this setting, we again find that the p^e-th roots and test ideals are monomial.

For all sufficiently large primes such that $p \equiv 1 \pmod{ab}$, we describe explicitly the chains of p^e-th roots, test ideals, and F-jumping numbers, and observe that they coincide for f and g. Finally, we note that the test ideals in this setting arise as reductions modulo p of the corresponding multiplier ideals.

2. Invariants of singularities in characteristic $p > 0$

Throughout, let R denote a ring of characteristic $p > 0$, and let $F\colon R \to R$, $f \mapsto f^p$, be the Frobenius endomorphism of R. For a nonnegative integer $e \geq 0$, the e-th iterated Frobenius is the endomorphism $F^e\colon R \to R$, $f \mapsto f^{p^e}$. In this section we introduce invariants of singularities in positive characteristic of interest to us. Often, these are referred to as F-invariants for they originate from the action of the Frobenius on R.

Restriction of scalars along F^e endows R with an exotic R-module structure denoted F^e_*R. Its elements are written as F^e_*x for $x \in R$. As an abelian group with respect to addition, F^e_*R is isomorphic to R. The action of R on F^e_*R is given by restriction of scalars: $r \cdot F^e_*x := F^e_*(r^{p^e}x)$, for $r \in R$, $F^e_*x \in F^e_*R$. A Noetherian ring R of characteristic $p > 0$ is said to be F-finite provided F^e_*R is a finite R-module for some $e \geq 1$ (equiv. all $e \geq 1$).

Example 2.1. Let $R = K[x_1, \dots, x_n]$ be a polynomial ring over a field K of characteristic $p > 0$. If K is perfect, i.e. the Frobenius $F\colon K \xrightarrow{\simeq} K$ is an automorphism of K, then F^e_*R splits as

$$F^e_*R \simeq \bigoplus_{0 \leq i_1, \dots, i_n < p^e} R\, F^e_*x_1^{i_1} \cdots x_n^{i_n},$$

therefore R is an F-finite ring, and F^e_*R is a finite free module with basis $\{F^e_*x_1^{i_1} \cdots x_n^{i_n} \mid 0 \leq i_1, \dots, i_n < p^e\}$. In the sequel, we will refer to this as the standard basis of F^e_*R.

2.1. Frobenius powers and p^e-th roots of ideals

Let R be a regular F-finite ring of characteristic $p > 0$.

Definition 2.2. Let I be an ideal of R. For an integer $e \geq 0$, the e-th Frobenius power of I is the ideal

$$I^{[p^e]} = F^e(I)R = (f^{p^e} \mid f \in I).$$

Remark 2.3. One checks that if $I = (f_\lambda \mid \lambda \in \Lambda)$ is a generating set for I, then $I^{[p^e]} = (f_\lambda^{p^e} \mid \lambda \in \Lambda)$, hence $I^{[p^e]} \subseteq I^{p^e}$. The reverse containment holds when $I = (f)$ is principal, that is, $(f)^{[p^e]} = (f)^{p^e}$, thus Frobenius powers and regular powers coincide for principal ideals.

A sort of "converse" operation to the Frobenius power is the ideal of p^e-th roots of an ideal I. These were introduced in [1] in the principal case under the notation $I_e(f)$, and later on exploited in [3] to give an alternative definition of the test ideals (see Section 2.2), using the notation $I^{[1/p^e]}$.

Definition 2.4. A Cartier operator of level $e \geq 0$ is an R-linear map $F_*^e R \to R$. The set of Cartier operators of level $e \geq 0$ has a natural R-module structure, which we denote by $\mathcal{C}_R^e := \mathrm{Hom}_R(F_*^e R, R)$.

Definition 2.5. Let I be an ideal of R. For an integer $e \geq 0$, the ideal of p^e-th roots of I is the ideal

$$\mathcal{C}_R^e \cdot I = (\varphi(F_*^e f) \mid \varphi \in \mathcal{C}_R^e,\ f \in I).$$

Remark 2.6. If R is a regular and F-finite, $\mathcal{C}_R^e \cdot I$ is characterized as the smallest ideal of R in the sense of inclusion such that $I \subseteq (\mathcal{C}_R^e \cdot I)^{[p^e]}$.

Proposition 2.7 ([1], [3, Proposition 2.5])**.** *Suppose that $F_*^e R$ is a free R-module with basis $F_*^e x_1, \dots, F_*^e x_n$. For an ideal $I = (f_1, \dots, f_m)$ of R, let*

$$F_*^e f_i = \sum_{1 \leq j \leq n} f_{ij} F_*^e x_j$$

be the expression in the basis of f_i, $i = 1, \dots, m$. Then $\mathcal{C}_R^e \cdot I = (f_{ij} \mid 1 \leq i \leq m,\ 1 \leq j \leq n)$.

We collect below a few facts about p^e-th roots that will be useful later on; for a proof, we refer the reader to [3, Lemma 2.4].

Lemma 2.8. *Let I, J be ideals of R, and $d, e \geq 0$ be nonnegative integers.*

(i) If $I \subseteq J$, then $\mathcal{C}_R^e \cdot I \subseteq \mathcal{C}_R^e \cdot J$.

(ii) One has that $J \cdot (\mathcal{C}_R^e \cdot I) = \mathcal{C}_R^e \cdot (I \cdot J^{[p^e]})$.

(iii) One has that $\mathcal{C}_R^e \cdot I = \mathcal{C}_R^{d+e} \cdot I^{[p^d]}$. In particular, if $I = (f)$ is principal, then $\mathcal{C}_R^e \cdot f = \mathcal{C}_R^{d+e} \cdot f^{p^d}$.

2.2. Test ideals, F-jumping numbers, and ν-invariants

Let R denote regular F-finite ring of characteristic $p > 0$. We now introduce the generalized test ideals of an ideal in R, as defined in [3], along with their associated invariants. Throughout, we denote by $\lceil x \rceil$ the ceiling of a real number.

Definition 2.9. Fix an ideal I of R. The test ideal of I with exponent $\lambda \in \mathbb{R}_{\geq 0}$ is

$$\tau(I^\lambda) = \bigcup_{e \geq 0} \mathcal{C}_R^e \cdot I^{\lceil \lambda p^e \rceil}.$$

Remark 2.10. It can be shown that the p^e-th roots appearing on the right-hand side give an ascending chain of ideals, which eventually stabilizes because R is Noetherian, therefore $\tau(I^\lambda) = \mathcal{C}_R^e \cdot I^{\lceil \lambda p^e \rceil}$ for $e \gg 0$.

Since p^e-th roots preserve inclusions (Lemma 2.8), so do test ideals, that is, $\tau(I^\lambda) \supseteq \tau(I^\mu)$ whenever $\lambda \leq \mu$. It follows the test ideals give a descending family of ideals obtained as λ ranges over the nonnegative real numbers. This chain is right semi-continuous in the following sense:

Theorem 2.11 ([9, Remark 2.12], [3, Corollary 2.16, Theorem 3.1])**.** *Let I be an ideal of R.*

(i) For each $\lambda \geq 0$, there exists $\varepsilon > 0$ such that $\tau(I^\lambda) = \tau(I^{\lambda+\varepsilon})$.

(ii) There exist real numbers $\lambda > 0$ such that $\tau(I^{\lambda-\varepsilon}) \supsetneq \tau(I^\lambda)$ for all $\varepsilon > 0$.

Definition 2.12. A positive real number $\lambda > 0$ is an F-jumping number of an ideal I of R provided

$$\tau(I^{\lambda-\varepsilon}) \supsetneq \tau(I^\lambda), \quad \text{for all } \varepsilon > 0.$$

The F-pure threshold of I, written $\operatorname{fpt}(I)$, is defined as the infimum among the F-jumping numbers of I. It is characterized by $\operatorname{fpt}(I) = \sup\{\lambda > 0 \mid \tau(I^\lambda) = R\}$.

Theorem 2.13 (Skoda's theorem, [3, Proposition 2.25])**.** *Suppose that I is an ideal generated by n elements. Then $\tau(I^\lambda) = I \cdot \tau(I^{\lambda-1})$ for every real number $\lambda \geq n$.*

The result below shows that the log-canonical threshold of a polynomial with integer coefficients can be recovered from the F-pure thresholds of the reductions modulo p:

Theorem 2.14 ([9, Theorem 3.4])**.** *Let $f \in \mathbb{C}[x_1, \dots, x_n]$ be a polynomial with rational coefficients. For a prime number $p > 0$, let $f_p \in \mathbb{F}_p[x_1, \dots, x_n]$ be the reduction of f modulo p. Then*

$$\operatorname{lct}(f) = \lim_{p\to\infty} \operatorname{fpt}(f_p).$$

It is conjectured that the statement above generalizes, in a sense, to multiplier and test ideals. Before announcing it, let us remark that if s is a Zariski-closed point in the spectrum of a finitely generated $\mathbb{Z}$-algebra A, that is, s is a maximal ideal of A, then the quotient A/sA has positive characteristic.

Conjecture 2.15 (Weak ordinarity conjecture, [8, Conjecture 1.2])**.** *Let I be an ideal in the polynomial ring $\mathbb{C}[x_1, \dots, x_n]$. Suppose that I can be generated by elements in a finitely generated $\mathbb{Z}$-subalgebra A of $\mathbb{C}[x_1, \dots, x_n]$. Then there exists a Zariski-dense subset S of* $\operatorname{Spec} A$ *consisting of closed points such that*

$$\mathcal{J}(I^\lambda)_s = \tau(I_s^\lambda), \quad \textit{for all } \lambda \geq 0,$$

for every $s \in S$, where I_s^λ, $\mathcal{J}(I^\lambda)_s$ denote the image under $A \to A/sA$ of I^λ, and $\mathcal{J}(I^\lambda)$, respectively.

Closely related to the F-jumping numbers are the F-thresholds introduced in [9]. Their definition is based on a different family of invariants, namely, the ν-invariants, which are of interest by themselves.

Definition 2.16. Let I, J be ideals of R such that $I \subseteq \sqrt{J}$, where $\sqrt{J}$ denotes the radical of J. The ν-invariant of level $e \geq 0$ of I with respect to J is

$$\nu_I^J(p^e) = \max\{n \geq 0 \mid I^n \subseteq J^{[p^e]}\}.$$

Let us denote by $\nu_I^\bullet(p^e)$ the set of ν-invariants of level $e \geq 0$ of I.

Definition 2.17. Let I, J be ideals of R such that $I \subseteq \sqrt{J}$. The F-threshold of I with respect to J is

$$c^J(I) = \lim_{e\to\infty} \frac{\nu_I^J(p^e)}{p^e}.$$

Theorem 2.18 ([3, Corollary 2.30, Theorem 3.1])**.** *The set of F-jumping numbers of an ideal I coincides with the set of F-thresholds $c^J(I)$ of I obtained as J ranges over all the ideals of R satisfying $I \subseteq \sqrt{J}$. In particular, both sets are rational and discrete.*

In computing F-thresholds, different ideals J, J' containing I in their radical may give raise to the same F-threshold. Instead, however, one can look at the spots where the chain of ideals below jumps:

$$\cdots \supseteq \mathcal{C}_R^e \cdot I^{r-1} \supseteq \mathcal{C}_R^e \cdot I^r \supseteq \mathcal{C}_R^e \cdot I^{r+1} \supseteq \cdots.$$

Definition 2.19 ([10, Proposition 4.2])**.** The set of ν-invariants of level $e \geq 0$ of an ideal I of R is

$$\nu_I^\bullet(p^e) = \{r \geq 0 \mid \mathcal{C}_R^e \cdot I^r \supsetneq \mathcal{C}_R^e \cdot I^{r+1}\}.$$

2.3. Chains of p^e-th roots and r-invariants

Through this section, let $R = K[x_1, \dots, x_d]$ denote a polynomial ring over a field K of characteristic $p > 0$. We are interested in obtaining the test ideals $\tau(f^\lambda)$ of a polynomial $f \in R$. Skoda's theorem (see Theorem 2.13) shows that $\tau(f^\lambda) = (f)\tau(f^{\lambda-1})$ for every real number $\lambda \geq 1$, hence it suffices to look at test ideals with $0 < \lambda < 1$. In this case, by Remark 2.10, one has that $\tau(f^\lambda) = \mathcal{C}_R^e \cdot f^r$ for some integer $r \leq p^e$. Aside from test ideals, we are keen on p^e-th roots $\mathcal{C}_R^e \cdot f^n$. By writing n uniquely as $n = sp^e + r$, with $s \geq 0$, $0 \leq r < p^e$, it follows from Lemma 2.8 that $\mathcal{C}_R^e \cdot f^n = (f)^s \cdot \mathcal{C}_R^e \cdot f^r$. Altogether, this shows it is enough to consider the ideals in chains of the form

$$R = \mathcal{C}_R^e \cdot f^0 \supseteq \mathcal{C}_R^e \cdot f \supseteq \mathcal{C}_R^e \cdot f^2 \supseteq \cdots \supseteq \mathcal{C}_R^e \cdot f^{p^e-2} \supseteq \mathcal{C}_R^e \cdot f^{p^e-1}. \tag{1}$$

Definition 2.20. We refer to (1) as the chain of ideals of p^e-th roots of f.

Lemma 2.21. *For a polynomial $f \in R$, one has that $\nu_f^\bullet(p^e) = (\nu_f^\bullet(p^e) \cap [0, p^e)) + p^e\mathbb{Z}_{\geq 0}$.*

Notation 2.22. *If $\underline{u} = (u_1, \dots, u_d) \in \mathbb{Z}_{\geq 0}^d$ is a multi-index, we let $\underline{x}^{\underline{u}}$ be the monomial $\underline{x}^{\underline{u}} = x_1^{u_1} \cdots x_d^{u_d}$.*

Perhaps, the most straightforward way to detect a jump $\mathcal{C}_R^e \cdot f^r \supsetneq \mathcal{C}_R^e \cdot f^{r+1}$ in chain (1) is to test if a monomial $\underline{x}^{\underline{u}}$ in $\mathcal{C}_R^e \cdot f^r$ drops from $\mathcal{C}_R^e \cdot f^{r+1}$, which means r is a ν-invariant attached to f and the monomial $\underline{x}^{\underline{u}}$. While this technique is standard in the field, to the best of our knowledge, the invariant "r" has not been assigned a name in the literature. We therefore introduce the following definition:

Definition 2.23. We define the r-invariant of level $e \geq 0$ of f with respect to a monomial $\underline{x}^{\underline{u}}$ by

$$r_R^e(f; \underline{x}^{\underline{u}}) = \sup\{n \in \mathbb{Z} \mid \underline{x}^{\underline{u}} \in \mathcal{C}_R^e \cdot f^n\}.$$

Lemma 2.24. *Let $f \in R$ be a polynomial. Suppose that f is not a monomial.*

(i) For all monomials $\underline{x}^{\underline{u}}$ in R, $0 \leq r_R^e(f, \underline{x}^{\underline{u}}) \leq p^e - 1$.

(ii) Every r-invariant of f is a ν-invariant.

When every ideal in the chain of p^e-th roots of f is monomial, the converse to Lemma 2.24 holds. In spite of how restrictive this latter condition may seem, we will come across it in Section 3.

Lemma 2.25. *Let $f \in R$ be a polynomial that is not a monomial. Suppose that every ideal in the chain of p^e-th roots of f is monomial.*

(i) One has that $\nu_f^\bullet(p^e) \cap [0, p^e) = \{r_R^e(f; \underline{x}^{\underline{u}}) \mid \underline{x}^{\underline{u}}\}$.

(ii) For an integer $0 \le n < p^e$, one has that $\mathcal{C}_R^e \cdot f^n = (\underline{x}^{\underline{u}} \mid r_R^e(f; \underline{x}^{\underline{u}}) \le n)$.

3. Ideals of p^e-th roots of plane curves

In this section, we describe the invariants previously introduced, for quasi-homogeneous plane curves defined over perfect fields of characteristic $p > 0$, for infinitely many primes, and their one-monomial constant Milnor number deformations. Throughout, let $\lfloor x \rfloor$ and $\lceil x \rceil$ be the floor and ceil functions, respectively.

Remark 3.1. Let $R = K[x_1, \dots, x_d]$ be a polynomial ring over a perfect field K of characteristic $p > 0$, so $F_*^e R$ is a free R-module with standard basis $\{F_*^e x_1^{i_1} \cdots x_d^{i_d} \mid 0 \le i_1, \dots, i_d < p^e\}$ (Example 2.1). Given a monomial $x_1^{u_1} \cdots x_d^{u_d}$, write each exponent uniquely as $u_i = s_i p^e + r_i$, with $s_i \ge 0$, $0 \le r_i < p^e$, for $i = 1, \dots, d$. Then

$$F_*^e(x_1^{u_1} \cdots x_n^{u_d}) = x_1^{s_1} \cdots x_n^{s_d} \, F_*^e(x_1^{r_1} \cdots x_n^{r_d})$$

is the basis expression of $x_1^{u_1} \cdots x_d^{u_d}$. Note that $s_i = \lfloor u_i / p^e \rfloor$, and r_i is the only integer $0 \le r_i < p^e$ with $u_i \equiv r_i \pmod{p^e}$. This calculation extends linearly to polynomials of R.

Definition 3.2. In the setting above, we say the monomials $x_1^{u_1} \cdots x_d^{u_d}$, $x_1^{v_1} \cdots x_d^{v_d}$ appear with the same basis element if $F_*^e x_1^{u_1} \cdots x_d^{u_d}$, $F_*^e x_1^{v_1} \cdots x_d^{v_d}$ lie in the same rank-one free R-submodule of $F_*^e R$ spanned by an element of the standard basis of $F_*^e R$. This is equivalent to $u_i \equiv v_i \pmod{p^e}$, for $i = 1, \dots, d$.

3.1. Ideals of p^e-th roots of quasi-homogeneous plane curves

Definition 3.3. Let K be a field. A quasi-homogeneous plane curve defined over K is the vanishing locus in K^2 of a polynomial of the form $f = x^a + y^b$, where $a, b \ge 2$. For simplicity, we will refer to the binomial $f = x^a + y^b$ as the quasi-homogeneous plane curve.

From now on, we work over the polynomial ring $R = K[x, y]$, with K perfect of characteristic $p > 0$. We remark, however, that all results remain valid upon relaxing the assumption on K to merely being F-finite, i.e. that K/K^{p^e} be a finite extension for some $e > 0$ (equiv. all $e > 0$).

Proposition 3.4. *Let $f = x^a + y^b$ be a quasi-homogeneous plane curve. Suppose that p does not divide a or b. For every integer $0 \le n < p^e$, one has that*

$$\mathcal{C}_R^e \cdot f^n = \left(x^{\lfloor ai/p^e \rfloor} y^{\lfloor bj/p^e \rfloor} \;\middle|\; i + j = n \text{ and } \binom{n}{i, j} \not\equiv 0 \pmod{p} \right).$$

In particular, every ideal in the chain of p^e-th roots of f is monomial.

Remark 3.5. In view of Proposition 3.4, $\mathcal{C}_R^e \cdot f^n$ contains the monomial $x^u y^v$, if and only if there exists a pair (i,j) of nonnegative integers such that:

$$\left\lfloor \frac{ai}{p^e} \right\rfloor \leq u, \quad \text{and} \quad \left\lfloor \frac{bj}{p^e} \right\rfloor \leq v, \quad \text{and} \quad \binom{n}{i,j} \not\equiv 0 \pmod p, \quad \text{and} \quad i+j=n.$$

The first two conditions are equivalent to $ai \leq (u+1)p^e - 1$, and $bj \leq (v+1)p^e - 1$, respectively. As a result, the r-invariant $r_R^e(f; x^u y^v)$ is the solution to the following linear integer programming problem:

$$\begin{cases} \text{maximize:} & i+j, \\ \text{subject to:} & ai \leq (u+1)p^e - 1, \\ & bj \leq (v+1)p^e - 1, \\ & \binom{i+j}{i,j} \not\equiv 0 \pmod p, \\ & i,j \in \mathbb{Z}_{\geq 0}. \end{cases} \tag{P1}$$

A solution $\geq p^e$ has no meaning by Lemma 2.24, hence it should be thought of as $x^u y^v \in \mathcal{C}_R^e \cdot f^{p^e - 1}$.

One can give general bounds for the optimum of (P1). To obtain a solution, however, it is helpful to make assumptions on the congruence class of p modulo ab. In what follows, we provide the solution under such additional assumption. Later on, in Section 3.3, we study the consequences on the F-invariants.

Lemma 3.6. *Let $f = x^a + y^b$ be a quasi-homogeneous plane curve. Suppose that $p \equiv 1 \pmod{ab}$. The r-invariant $r_R^e(f; x^u y^v)$ of a monomial $x^u y^v$ is*

$$r_R^e(f; x^u y^v) = \begin{cases} \left(\frac{u+1}{a} + \frac{v+1}{b}\right)(p^e - 1) & \text{if } bu + av < ab - a - b, \\ p^e - 1 & \text{if } bu + av \geq ab - a - b. \end{cases}$$

3.2. One-monomial deformations of quasi-homogeneous plane curves

Let $C \subseteq K^2$ be a plane curve given as the zero locus of a polynomial $h \in K[x,y]$. Suppose C passes through the origin. Then the Milnor number of C is defined as $\mu = \dim_K K[x,y]/(\partial_x h, \partial_y h)$. When $K = \mathbb{C}$, and h is a quasi-homogeneous plane curve $x^a + y^b$, the Milnor number determines the topological type of the singularity of C at the origin under deformations, in the following sense:

Theorem 3.7 ([7]). *Let $f = x^a + y^b$ be a quasi-homogeneous plane curve defined over $\mathbb{C}$, and g a deformation $g = x^a + y^b + t_1 x^{\alpha_1} y^{\beta_1} + \cdots + t_n x^{\alpha_n} y^{\beta_n}$, with $t_1, \ldots, t_n \in \mathbb{C}$. Suppose every deformation monomial $x^{\alpha_i} y^{\beta_i}$ on which g is supported (i.e. $t_i \neq 0$) satisfies $0 \leq \alpha_i < a-1$, $0 \leq \beta_i < b-1$, and $a\beta_i + b\alpha_i > ab$. Then f and g have the same Milnor number.*

Definition 3.8. A constant Milnor number deformation, or μ-constant deformation of a quasi-homogeneous plane curve $f = x^a + y^b$ defined over K, is the vanishing locus in K^2 of a polynomial of the form

$$g = x^a + y^b + \sum_{i=1}^{n} t_i x^{\alpha_i} y^{\beta_i}, \quad \text{where } t_i \in K,$$

with $0 \leq \alpha_i < a-1$, $0 \leq \beta_i < b-1$, and $a\beta_i + b\alpha_i > ab$, for $i = 1, \ldots, n$.

Hereinafter, we consider one-monomial μ-constant deformations of quasi-homogeneous plane curves defined over a perfect field K of characteristic $p > 0$, thus we work over the polynomial ring $R = K[x, y]$.

Proposition 3.9. *Let $g = x^a + y^b + tx^\alpha y^\beta$ be a μ-constant deformation of $f = x^a + y^b$. Suppose that p does not divide a or b, and $p > a\beta + b\alpha - ab$. For every integer $0 \leq n < p^e$, one has that*

$$\mathcal{C}_R^e \cdot g^n = \left(x^{\lfloor (ai+\alpha k)/p^e \rfloor} y^{\lfloor (bj+\beta k)/p^e \rfloor} \;\middle|\; i + j + k = n \text{ and } \binom{n}{i, j, k} \not\equiv 0 \pmod{p} \right).$$

Proposition 3.10. *Let $g = x^a + y^b + tx^\alpha y^\beta$ be a μ-constant deformation of the quasi-homogeneous plane curve $f = x^a + y^b$. Suppose that p does not divide a or b, and $p > a\beta + b\alpha - ab$. One has that $\mathcal{C}_R^e \cdot f^n \subseteq \mathcal{C}_R^e \cdot g^n$ for every integer $0 \leq n < p^e$.*

Remark 3.11. By Proposition 3.9, given an integer $0 \leq n < p^e$, a monomial $x^u y^v$ is in $\mathcal{C}_R^e \cdot g^n$ if and only if there exists a triple (i, j, k) of nonnegative integers such that:

$$\left\lfloor \frac{ai + \alpha k}{p^e} \right\rfloor \leq u, \quad \text{and} \quad \left\lfloor \frac{bj + \beta k}{p^e} \right\rfloor \leq v, \quad \text{and} \quad \binom{n}{i, j, k} \not\equiv 0 \pmod{p}, \quad \text{and} \quad i + j + k = n.$$

One sees the first two conditions are equivalent to $ai + \alpha k \leq (u+1)p^e - 1$, and $bj + \beta k \leq (v+1)p^e - 1$. It follows that $r_R^e(g; x^u y^v)$ is the solution to the following linear integer programming problem:

$$\begin{cases} \text{maximize:} & i + j + k, \\ \text{subject to:} & ai + \alpha k \leq (u+1)p^e - 1, \\ & bj + \beta k \leq (v+1)p^e - 1, \\ & \binom{i+j+k}{i,j,k} \not\equiv 0 \pmod{p}, \\ & i, j, k \in \mathbb{Z}_{\geq 0}. \end{cases} \tag{P2}$$

By Proposition 3.10, a solution of (P2) is bounded below by a solution of (P1). As in Remark 3.5, a solution $\geq p^e$ must be thought of as $x^u y^v \in \mathcal{C}_R^e \cdot g^{p^e - 1}$.

Lemma 3.12. *Let $g = x^a + y^b + tx^\alpha y^\beta$ be a μ-constant deformation of a quasi-homogeneous plane curve $f = x^a + y^b$. Suppose that $p \equiv 1 \pmod{ab}$. For a monomial $x^u y^v$, one has that*

$$r_R^e(f; x^u y^v) = r_R^e(g; x^u y^v).$$

3.3. *F*-invariants of quasi-homogeneous plane curves and deformations

To conclude, we use the r-invariants of quasi-homogeneous plane curves and their one-monomial μ-constant deformations to compute their F-invariants when $p \equiv 1 \pmod{ab}$.

Proposition 3.13. *Let h be either the quasi-homogeneous plane curve $f = x^a + y^b$, or the μ-constant deformation $g = x^a + y^b + tx^\alpha y^\beta$. Suppose that $p \equiv 1 \pmod{ab}$.*

(i) The p^e-th root of h^n, with $0 \le n < p^e$, is

$$\mathcal{C}_R^e \cdot h^n = \left(x^u y^v \;\middle|\; \frac{u+1}{a} + \frac{v+1}{b} \le \frac{n}{p^e - 1} \right).$$

(ii) The ν-invariants of h of level e are

$$\nu_h^\bullet(p^e) = \left\{ kp^e + \left(\frac{u+1}{a} + \frac{v+1}{b} \right)(p^e - 1),\ (k+1)p^e - 1 \;\middle|\; bu + av < ab - a - b,\ k \ge 0 \right\}.$$

Theorem 3.14. *Let h be either the quasi-homogeneous plane curve $f = x^a + y^b$, or the μ-constant deformation $g = x^a + y^b + tx^\alpha y^\beta$. Suppose that $p \equiv 1 \pmod{ab}$, and $p > a\beta + b\alpha - ab$.*

(i) The F-jumping numbers of h are

$$\mathrm{FJN}(h) = \left\{ \frac{u+1}{a} + \frac{v+1}{b}, 1 \;\middle|\; bu + av < ab - a - b \right\} + \mathbb{Z}_{\ge 0}.$$

(ii) The test ideal of h with exponent $\lambda \in (0, 1)$ is

$$\tau(h^\lambda) = \left(x^u y^v \;\middle|\; \frac{u+1}{a} + \frac{v+1}{b} > \lambda \right).$$

Remark 3.15. For a prime number p sufficiently large with $p \equiv 1 \pmod{ab}$, Proposition 3.13 and Theorem 3.14 show that a quasi-homogenous plane curve and a one-monomial μ-constant deformation have the same chains of p^e-th roots, ν-invariants, and F-jumping numbers. Furthermore, their test ideals coincide for $\lambda \in (0, 1)$.

Remark 3.16. Let $f = x^a + y^b$ be a quasi-homogeneous plane curve, or a one-monomial μ-constant deformation $g = x^a + y^b + tx^\alpha y^\beta$, $t \in \mathbb{Z}$, defined over $\mathbb{C}$. Consider the sets $\mathcal{X}_f = \{p \mid p \equiv 1 \pmod{ab},\ p \text{ prime}\}$, $\mathcal{X}_g = \{p \mid p \equiv 1 \pmod{ab},\ p > a\beta + b\alpha - ab,\ p \text{ prime}\}$ as subspaces of $\operatorname{Spec}\mathbb{Z}$, which are infinite by Dirichlet's theorem on arithmetic progressions. In the Zariski topology on $\mathbb{Z}$, a nonempty open subset is the complement of the union of finitely many points, and hence must intersects both $\mathcal{X}_f$ and $\mathcal{X}_g$. It follows that these sets are dense in $\operatorname{Spec}\mathbb{Z}$.

Choose a prime $p \in \mathcal{X}_f$, and denote by f_p the reduction modulo p of f along $\mathbb{Z}[x, y] \to \mathbb{F}_p[x, y]$. Similarly, let g_p be the reduction of g, with $p \in \mathcal{X}_g$. The multiplier ideals $\mathcal{J}(f^\lambda)$, $\mathcal{J}(g^\lambda)$ are generated by polynomials over $\mathbb{Z}$, and can be obtained with an algorithm proposed by Blanco and Dachs-Cadefau [2]. After computing their reductions $\mathcal{J}(f^\lambda)_p$, $\mathcal{J}(g^\lambda)_p$, one sees they coincide with the test ideals, namely $\mathcal{J}(f^\lambda)_p = \tau(f_p^\lambda)$, and $\mathcal{J}(g^\lambda)_p = \tau(g_p^\lambda)$, for all λ. In consequence, the weak ordinarity conjecture (Conjecture 2.15) holds for quasi-homogeneous plane curves and their one-monomial μ-constant deformations.

Acknowledgements

The author would like to express his sincere gratitude to his advisor, Josep Àlvarez Montaner, and to the master's thesis committee for evaluating the project from which this work originates. He also wishes to thank the Institut d'Estudis Catalans and, in particular, the Societat Catalana de Matemàtiques and the jury members for awarding the Évariste Galois Prize to the master's thesis project. Finally, the author is grateful to the editors and the anonymous referee for their valuable feedback and suggestions.

References

[1] J. Àlvarez-Montaner, M. Blickle, G. Lyubeznik, Generators of D-modules in positive characteristic, *Math. Res. Lett.* **12(4)** (2005), 459–473.

[2] G. Blanco, F. Dachs-Cadefau, Computing multiplier ideals in smooth surfaces, in: *Extended Abstracts February 2016—Positivity and Valuations*, Trends Math. Res. Perspect. CRM Barc. **9**, Birkhäuser/Springer, Cham, 2018, pp. 57–63.

[3] M. Blickle, M. Mustaţă, K.E. Smith, Discreteness and rationality of F-thresholds, Special volume in honor of Melvin Hochster, *Michigan Math. J.* **57** (2008), 43–61.

[4] N. Hara, Ken-Ichi Yoshida, A generalization of tight closure and multiplier ideals, *Trans. Amer. Math. Soc.* **355(8)** (2003), 3143–3174.

[5] M. Hochster, C. Huneke, Tight closure, invariant theory, and the Briançon–Skoda theorem, *J. Amer. Math. Soc.* **3(1)** (1990), 31–116.

[6] R. Lazarsfeld, *Positivity in Algebraic Geometry II. Positivity for Vector Bundles, and Multiplier Ideals*, Ergeb. Math. Grenzgeb. (3) **49** [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics], Springer-Verlag, Berlin, 2004.

[7] Lê Dũng Tráng, C.P. Ramanujam, The invariance of Milnor's number implies the invariance of the topological type, *Amer. J. Math.* **98(1)** (1976), 67–78.

[8] M. Mustaţă, V. Srinivas, Ordinary varieties and the comparison between multiplier ideals and test ideals, *Nagoya Math. J.* **204** (2011), 125–157.

[9] M. Mustaţă, S. Takagi, Kei-ichi Watanabe, F-thresholds and Bernstein–Sato polynomials, in: *European Congress of Mathematics*, European Mathematical Society (EMS), Zürich, 2005, pp. 341–364.

[10] E. Quinlan-Gallego, Bernstein–Sato theory for arbitrary ideals in positive characteristic, *Trans. Amer. Math. Soc.* **374(3)** (2021), 1623–1660.

REPORTS@SCM
AN ELECTRONIC JOURNAL OF THE
SOCIETAT CATALANA DE MATEMÀTIQUES

Use of Fourier series in $\mathbb{S}^2$ to approximate star-shaped surfaces

Miguel Nasarre Budiño
Universitat Autònoma
de Barcelona
miguel.nasarre.budino@gmail.com

Resum *(CAT)*
Primer ens centrem a estendre la noció de sèrie de Fourier, estudiant com podem representar funcions de quadrat integrable sobre varietats de Riemann. Per fer això ens ajudem del teorema de Hodge, que ens permetrà trobar bases d'aquests espais a partir del laplacià.
Després veiem com aquest mètode es pot utilitzar per trobar les sèries de Fourier per a funcions periòdiques, i per a les funcions $L^2(\mathbb{S}^2)$, el cas principal. També discutim com estimar l'error en norma L^2, i implementem totes les fórmules que es troben a l'article a un programa per poder visualitzar els resultats obtinguts.

Abstract *(ENG)*
First we will focus in extending the notion of Fourier series, studying how can we represent functions of integrable square over Riemannian manifolds. To do this we will use the Hodge theorem, that will allow us to find basis of these spaces through the Laplacian.
Then we will see how this method can be used to find the Fourier series for periodic functions, and for the functions $L^2(\mathbb{S}^2)$, our main case of study. We will also discuss how to estimate the L^2-error, and we implement all the formulas found in the article in a program to be able to visualize the obtained results.

Keywords: *Laplacian, Riemannian manifold, basis, spherical harmonics, Fourier series, L^2-error estimates.*
MSC (2020): *Primary 55M25, 57P10. Secondary 55P15, 57R19, 57N15.*

Received: *July 29, 2025.*
Accepted: *August 25, 2025.*

Societat Catalana de Matemàtiques
Institut d'Estudis Catalans

1. Introduction

This article will follow the steps that were set in my bachelor's thesis and explain how to build Fourier series to be able to represent star-shaped surfaces.

In particular we will look at the construction of the Laplacian operator Δ induced by a metric g, on a set of functions $L^2(X)$, where X is a Riemannian manifold, and how we can use the Hodge theorem to look for basis of the function space $\{\psi_i\}_{i\in\Lambda} \in L^2(X)$ in the eigenvectors of the Laplacian $\Delta\psi_i = \lambda_i\psi_i$.

We will primarily focus our attention on the case for $\mathbb{S}^2$, where such basis will be given by the spherical harmonics, and we will study how to apply these formulas for computational use, and also how to a priori estimate the L^2-error for a certain amount of coefficients.

All of this culminating in a desktop application, that given a star-shaped surface triangulation will use the formulas explained on the article to represent that shape as a Fourier series, proving the applicability of the mathematics explored on my thesis.

2. Construction of the Laplacian on a Riemannian manifold

Given (X, g) a compact Riemannian manifold, for further calculations we will assume

$$\begin{aligned} \exists \Phi \colon U \subset \mathbb{R}^n &\longrightarrow X \\ (x_1, \dots, x_n) &\longrightarrow \Phi(x_1, \dots, x_n) \end{aligned}$$

a parameterization with a dense image, $\overline{\Phi(U)} = X$.

Definition 2.1. Given the prior manifold we define the Hilbert space

$$L^2(X) = \{f : X \longrightarrow \mathbb{R},\ f \text{ integrable square}\}$$

with the scalar product for $f, h \in C^\infty(X)$,

$$\langle f, h\rangle = \int_X fh\, dV_g,$$

where dV_g is the differential induced by the metric.

To build the Laplacian we will work with the dense subset $C^\infty(X) \subset L^2(X)$.

Definition 2.2. Let $\Omega^0(X) = C^\infty(X)$ be the space of differential 0-forms of X. We define

$$\Omega^1(X) = \{\omega = f_1\, dx_1 + \cdots + f_n\, dx_n \mid f_i \in \Omega^0(X)\}$$

the space of differential 1-forms of X.

Definition 2.3. We define the differential operator d as

$$\begin{aligned} d\colon \Omega^0(X) &\longrightarrow \Omega^1(X) \\ f &\longrightarrow \frac{\partial f}{\partial x_1}\,dx_1 + \cdots + \frac{\partial f}{\partial x_n}\,dx_n. \end{aligned}$$

We intend to use the differential operator to build the Laplacian as a self-adjoint operator, $\Delta = d^* \circ d$. Therefore we need to also build a Hilbert space with a well defined scalar product on $\Omega^1(X)$.

Definition 2.4. Let $L^2(X) = A^0(X)$, then we define

$$A^1(X) = \{\omega = f_1\,dx_1 + \cdots + f_n\,dx_n \mid f_i \in A^0(X)\}$$

the set of L^2 1-forms of X, and we see $\overline{\Omega^1(X)} = A^1(X)$.

Now we want to build the tensor $g^*\colon T^*X \oplus T^*X \to \mathbb{R}$.

Let $e_1, \dots, e_n \in T_xX$ be an orthonormal basis for g, then the dual basis of the 1-forms $\omega_1, \dots, \omega_n \in T_x^*X$ so that $\omega_i(e_j) = \delta_{ij}$ will be an orthonormal basis for g^* of T^*X. Through this basis g^* is well defined.

Definition 2.5. With what we have seen so far we can introduce then the scalar product in $A^1(X)$ with this tensor. Given $\omega, \eta \in \Omega^1(X)$,

$$\langle \omega, \eta \rangle_{\Omega^1} = \int_X g_x^*(\omega(x), \eta(x))\,dV_g.$$

This scalar product satisfies the properties of a Hilbert space. Therefore, we have another Hilbert space in $A^1(X)$.

Now we can use the definition of the adjoint operator to find the adjoint of the differential, and finally build the Laplacian operator.

$$\langle df, \omega \rangle_{A^1} = \langle f, d^*\omega \rangle_{A^0}, \quad \forall f \in \Omega^0(X),\ \forall \omega \in \Omega^1(X),$$
$$\Delta = d^* \circ d.$$

3. The Hodge theorem

Now that we have a self-adjoint operator inside a Hilbert space, we will be using the Hodge theorem, specific for this case, to back up our claim that we can find basis for this function space using the Laplacian.

In this article we will just announce the theorem without further proof. For a proper proof as well as more in depth information please refer to [5, p. 32], where the Hodge theorem is established and proven.

Theorem 3.1. *Let (M,g) be a compact Riemannian manifold, oriented and connected. There exists an orthonormal basis for $L^2(M, g)$ that consists of eigenvectors of the Laplacian. All the eigenvalues are real and positive, except for 0 which is an eigenvalue of multiplicity 1. Every eigenvalue has finite multiplicity, and they only accumulate at infinity.*

Once we find those eigenvectors $\{\psi_i\}_{i\in\Lambda} \in L^2(X)$ that form a basis, we can write any given element $f \in L^2(X)$ as its Fourier series

$$f = \sum_{i\in\Lambda} \alpha_i \psi_i, \tag{1}$$

where the coefficients are the ones obtained through the scalar product

$$\alpha_i = \langle f, \psi_i \rangle. \tag{2}$$

4. Construction of the common Fourier series

To begin with the examples let us focus on the case for $X = \mathbb{S}^1$, with the parameterization

$$\Phi\colon (0, 2\pi) \longrightarrow \mathbb{S}^1 \subset \mathbb{R}^2$$
$$x \longrightarrow (\cos x, \sin x).$$

We will be using the metric $g = \frac{dx \otimes dx}{4\pi^2}$, giving us the following scalar product, given $f, h \in A^0(\mathbb{S}^1)$,

$$\langle f, h\rangle_{A^0} = \frac{1}{2\pi}\int_0^{2\pi} f(x)h(x)\,dx.$$

And following the definition for g^* we obtain the scalar product for $A^1(\mathbb{S}^1)$. Given $\omega, \eta \in A^1(\mathbb{S}^1)$, with $\omega(x) = f(x)\,dx$ and $\eta(x) = h(x)\,dx$,

$$\langle \omega, \eta\rangle_{A^1} = \frac{1}{2\pi}\int_0^{2\pi} g_x^*(f(x)\,dx, h(x)\,dx)\,dx = 2\pi\int_0^{2\pi} f(x)h(x)\,dx.$$

Then applying the definition of the adjoint operator we obtain, given any $f \in \Omega^0(\mathbb{S}^1)$ and any $\omega \in \Omega^1(\mathbb{S}^1)$ with $\omega = h(x)\,dx$,

$$\langle df, \omega\rangle_{A^1} = \langle f, d^*\omega\rangle_{A^0},$$
$$2\pi\int_0^{2\pi} f'(x)h(x)\,dx = \frac{1}{2\pi}\int_0^{2\pi} f(x)(d^*\omega)(x)\,dx.$$

Integrating by parts and after some algebra we end up obtaining the expression for the Laplacian

$$\Delta = -4\pi^2\partial_x^2.$$

Now we want to look for the eigenvectors of this operator. Therefore we want to obtain a set of orthonormal functions $\{\psi_i\}_{i\in\Lambda} \in L^2(\mathbb{S}^1)$ that satisfy the equation $\Delta\psi_i = \lambda_i\psi_i$. If we develop the prior expression, we see

$$\Delta\psi = \lambda\psi,$$
$$-4\pi^2\partial_x^2\psi(x) = \lambda\psi(x),$$
$$4\pi^2\psi'' = -\lambda\psi,$$

giving us the very natural orthonormal eigenvectors

$$\psi_0(x) = 1,$$
$$\psi_n(x) = \sqrt{2}\cos(nx),$$
$$\phi_n(x) = \sqrt{2}\sin(nx).$$

As we can easily prove by solving the differential equation, these are all the solutions despite lineal combinations of them. Therefore, now as described in the prior section we can create the Fourier series using this basis, giving us the common example for real Fourier series we all know. Given any $f \in L^2(\mathbb{S}^1)$, we can write it as

$$\begin{aligned} f(x) &= A_0 + \sum_{n=1}^{\infty}(A_n \cos(nx) + B_n \sin(nx)), \quad x \in (0, 2\pi), \\ A_0 &= \frac{1}{2\pi}\int_0^{2\pi} f(x)\, dx, \\ A_n &= \frac{1}{\pi}\int_0^{2\pi} f(x)\cos(nx)\, dx, \\ B_n &= \frac{1}{\pi}\int_0^{2\pi} f(x)\sin(nx)\, dx. \end{aligned}$$

We can also consider the case for $L^2(\mathbb{S}^1, \mathbb{C})$. In this space almost all the process is exactly the same, and we obtain the same self-adjoint operator. The only difference is that when defining the scalar product, this one requires a symmetry by the conjugate, therefore we define given $f, h \in L^2(\mathbb{S}^1, \mathbb{C})$,

$$\langle f, h\rangle = \frac{1}{2\pi}\int_0^{2\pi} f(x)\overline{h(x)}\, dx.$$

For this particular example $C^\infty(\mathbb{S}^1, \mathbb{C})$, the most common basis will be

$$\psi_n(x) = e^{inx}, \quad \forall n \in \mathbb{Z} \text{ with } \lambda_n = 4\pi^2 n^2.$$

Finally, given any $f \in L^2(\mathbb{S}^1, \mathbb{C})$, we can write it as

$$\begin{aligned} f(x) &= \sum_{n=-\infty}^{\infty} \alpha_n e^{inx}, \\ \alpha_n &= \frac{1}{2\pi}\int_0^{2\pi} f(x)e^{-inx}\, dx. \end{aligned}$$

5. Case for $L^2(\mathbb{S}^2)$, the spherical harmonics

To start with this case we will also start defining the parameterization,

$$\begin{aligned} \Phi\colon (0, 2\pi) \times (0, \pi) &\longrightarrow \mathbb{S}^2 \subset \mathbb{R}^3 \\ (\varphi, \theta) &\longrightarrow (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta). \end{aligned}$$

Through this parameterization we can find a Riemannian metric induced from $\mathbb{R}^3$, giving us

$$g = \sin^2\theta\, d\varphi \otimes d\varphi + d\theta \otimes d\theta.$$

Therefore the scalar product will be

$$\langle f, h\rangle_{A^0} = \int_{\mathbb{S}^2} fh\, dV_g = \int_0^{\pi}\int_0^{2\pi} f(\varphi, \theta)h(\varphi, \theta)\sin\theta\, d\varphi\, d\theta.$$

And following the definition for g^* we obtain the scalar product for $A^1(\mathbb{S}^2)$, given

$$\omega(\varphi,\theta) = f_1(\varphi,\theta)\, d\varphi + f_2(\varphi,\theta)\, d\theta,$$
$$\eta(\varphi,\theta) = h_1(\varphi,\theta)\, d\varphi + h_2(\varphi,\theta)\, d\theta,$$

we obtain

$$\langle \omega, \eta \rangle_{A^1} = \int_{\mathbb{S}^2} g^*(\omega,\eta)\, dV_g = \iint \left(\frac{f_1 h_1}{\sin\theta} + f_2 h_2 \sin\theta \right) d\varphi\, d\theta.$$

Applying the definition of the adjoint operator we obtain, given $f \in \Omega^0(\mathbb{S}^2)$ and $\omega \in \Omega^1(\mathbb{S}^2)$,

$$\langle df, \omega \rangle_{A^1} = \langle f, d^*\omega \rangle_{A^0}.$$

After some calculations we end up obtaining the following expression for the Laplacian,

$$\Delta = -\left(\frac{\partial_\varphi^2}{\sin^2\theta} + \frac{\partial_\theta(\sin\theta\, \partial_\theta)}{\sin\theta} \right).$$

At this point we introduce the spherical harmonics, which are functions with the following expression.

Definition 5.1. Let $\ell \in \mathbb{N}\cup\{0\}$ and $m \in \mathbb{Z}$ with $|m| \leq \ell$. We define the function $Y_\ell^m : \mathbb{S}^2 \to \mathbb{R}$ parameterized with the coordinates we have been using so far as

$$Y_\ell^m(\varphi,\theta) = \begin{cases} \sqrt{\dfrac{(2\ell+1)(\ell-m)!}{2\pi(\ell+m)!}} P_\ell^m(\cos\theta)\cos(m\varphi) & \text{if } m>0, \\ \sqrt{\dfrac{2\ell+1}{4\pi}} P_\ell^0(\cos\theta) & \text{if } m=0, \\ \sqrt{\dfrac{(2\ell+1)(\ell-|m|)!}{2\pi(\ell+|m|)!}} P_\ell^{|m|}(\cos\theta)\sin(|m|\varphi) & \text{if } m<0, \end{cases} \tag{3}$$

where P_ℓ^m are the associated Legendre polynomials. They are defined as the canonical solutions of the general Legendre equation

$$(1-x^2)\frac{d^2}{dx^2}P_\ell^m(x) - 2x\frac{d}{dx}P_\ell^m(x) + \left[\ell(\ell+1) - \frac{m^2}{1-x^2}\right] P_\ell^m(x) = 0. \tag{4}$$

These functions are the orthonormal basis obtained by the eigenvectors of the Laplacian. Unfortunately given the length constrains of this article the proper proof of such claim is not possible. For the interested reader I strongly recommend checking the full thesis for a very interesting proof as well as a more in depth explanation of all the steps we have done before and all the results we have seen.

Nevertheless I will try to give a somewhat satisfying overview of the proof found in the thesis. First we start by proving that the spherical harmonics are orthonormal; for this and further steps ahead we use [4, Chapter 14], [1, pp. 331–341], where a lot of the properties of the Legendre polynomials are found as well as some recurrences useful for their computation. In this case we use the Legendre polynomials orthogonality,

$$\int_{-1}^{1} P_\ell^m(x) P_k^m(x)\, dx = \frac{2(\ell+m)!}{(2\ell+1)(\ell-m)!}\delta_{\ell k}.$$

Which finally ends up helping us find the orthonormality result

$$\langle Y_\ell^m, Y_k^n \rangle = \delta_{\ell k}\delta_{mn}.$$

Then we evaluate the Laplacian of the spherical harmonics. To simplify the expression we use the substitution $v = \cos\theta$, where $v' = -\sin\theta = -\sqrt{1-v^2}$; with some rearrangement we can use (4) to get rid of one of the derivatives, allowing us to simplify the expression and with some algebra we obtain $\Delta Y_\ell^m = \ell(\ell+1)Y_\ell^m$, which proves that the spherical harmonics are eigenvectors of the Laplacian with eigenvalues $\lambda_\ell = \ell(\ell+1)$.

Finally we have to prove that these eigenvalues are sufficient to make a basis for $L^2(\mathbb{S}^2)$. For this final step we inspire on the beautiful proof found in Chapter 7 of the book *Groupes et symétries* [2]. We begin with the Stone–Weierstrass theorem.

Theorem 5.2. *Let X be a compact Hausdorff space and A a sub-algebra of the space of continuous functions from X to the real numbers $C^0(X)$, that has a constant non-zero function. Then A is dense in $C^0(X)$ if and only if it separates points.*

This tells us that the polynomials are dense for $C^0(X)$ as long as X is a compact Hausdorff space. Therefore the proof centers itself in proving that the harmonic polynomials are a basis for the restriction to $\mathbb{S}^2$ of all polynomials in $\mathbb{R}^3$, more specifically there are a total of $2\ell+1$ homogeneous harmonic polynomials of degree ℓ, and all together they form a basis of the polynomials in $\mathbb{R}^3$ restricted to the sphere. And therefore applying the Stone–Weierstrass theorem they are a basis for all continuous functions restricted to the sphere.

The second part of the proof uses the results we obtained earlier from the spherical harmonics to prove that these are indeed the restriction to $\mathbb{S}^2$ of homogeneous harmonic polynomials of degree ℓ, and since we have $2\ell+1$ spherical harmonics for every ℓ value, we conclude that these are a basis for $L^2(\mathbb{S}^2)$.

6. A priori estimations of the L^2-error

It would be interesting for further applications of the formulas we have seen so far if we had an a priori estimate of how many coefficients we need to compute to obtain a desired relative L^2-error. For a subset of frequencies $L \subset \Lambda$ we define this error as

$$\epsilon = \frac{||f - f^L||}{||f||},$$

where $f^L = \sum_{i \in L} \langle f, \psi_i \rangle \psi_i$.

To do this we will be using the methods found in [3]. This paper was developed by my tutor and is the initial inspiration behind the thesis, for proof of the theorems that will follow I strongly recommend looking at their article, or it can also be found in the thesis.

Theorem 6.1. *Let $f \in A^0(X)$ be a function so that $df \in A^1(X)$ and $\Delta f \in A^0(X)$ are well defined. For every $\epsilon > 0$ exists a finite subset $L_f(\epsilon) \subset \Lambda$ that only depends of $||f||$, $||df||$, $||\Delta f||$ and ϵ so that*

$$||f - f^{L_f(\epsilon)}|| \leq \epsilon ||f||. \tag{5}$$

Actually, $L_f(\epsilon)$ can be chosen by the preimage $\lambda \colon \Lambda \to \mathbb{R}$ of the compact interval $[L_f^-(\epsilon), L_f^+(\epsilon)]$, where

$$L_f^{\pm}(\epsilon) = \frac{||df||^2 \pm \epsilon^{-1}\sqrt{||\Delta f||^2||f||^2 - ||df||^4}}{||f||^2}. \tag{6}$$

Theorem 6.2. *Assume we already computed the Fourier coefficients of f for a given subset $I \subset \Lambda$. Then the inequality (5) is also satisfied for the new subset of coefficients*

$$L_f(\epsilon, I) = I \cup L_{f^{\Lambda \setminus I}}\left(\frac{\epsilon||f||}{||f^{\Lambda \setminus I}||}\right). \tag{7}$$

These formulas will be used in the programs that calculate and display the Fourier coefficients, in order to see their performance on a real scenario, and the results will be discussed at the end.

7. Formulas for computational use

In this section we will explain the formulas implemented in the program to calculate the Fourier series for the case $L^2(\mathbb{S}^2)$. For our particular case we will suppose that we have a triangulation $\{T_i\}_{i=0}^N$ of a star-shaped surface S of strictly positive radius. We will express this surface by its radius $r\colon \mathbb{S}^2 \to \mathbb{R}^+$ given any point of the sphere. To compute the Fourier coefficients for this function we will need to apply (2), and the expression will be as follows,

$$r_\ell^m = \int_{\mathbb{S}^2} r\, Y_\ell^m \, dV_g = \sum_{i=0}^N \int_{T_i^p} r\, Y_\ell^m \, dV_g,$$

where $T_i^p \subset \mathbb{S}^2$ is the triangle T_i projected onto the unit sphere. If the triangulation is fine enough, we can approximate that Y_ℓ^m is constant over the entire surface of the triangle, and since the mean radius of the triangle is the radius in the barycenter, we will rewrite the prior expression as

$$r_\ell^m \approx \sum_{i=0}^N ||\overline{T}_i||\, Y_\ell^m(\overline{T}_i^p) A(T_i^p),$$

where $\overline{T}_i$ is the barycenter of the triangle T_i, and $A(T_i^p)$ is the area of the spherical triangle T_i^p.

Now we will look at how to evaluate every term of this formula. Starting with $||\overline{T}||$, this is just the norm of the mean of the three $v_0, v_1, v_2 \in \mathbb{R}^3$ points that make the triangle, $||\overline{T}|| = ||\frac{v_0+v_1+v_2}{3}||$.

Then to calculate the area of the spherical triangle $A(T^p)$ we will use the formula $A = \alpha_0 + \alpha_1 + \alpha_2 - \pi$, where α_i are the angles of the spherical triangle. To find these angles we can use the tangent lines to the sphere $u = \frac{(v_i \times v_j)\times v_i}{||(v_i \times v_j)\times v_i||}$, and find the angles through the scalar product between them.

Finally we have to compute the spherical harmonic at the barycenter of the projected triangle. In this case we will explain how the program computes (3) for any given point $(x, y, z) \in \mathbb{S}^2 \subset \mathbb{R}^3$. To compute the trigonometric functions $\cos(m\varphi)$ and $\sin(m\varphi)$ we will use the fact that we know $\cos\varphi = \frac{x}{\sqrt{1-z^2}}$ and $\sin\varphi = \frac{y}{\sqrt{1-z^2}}$ and the Chebyshev polynomial, that give us

$$T_n(\cos\theta) = \cos(n\theta),$$
$$U_{n-1}(\cos\theta)\sin\theta = \sin(n\theta).$$

To compute them we will use the following recurrences,

$$\begin{aligned} T_0(x) &= 1, \\ T_1(x) &= x, \\ T_{n+1}(x) &= 2xT_n(x) - T_{n-1}(x), \\ U_0(x) &= 1, \\ U_1(x) &= 2x, \\ U_{n+1}(x) &= 2xU_n(x) - U_{n-1}(x). \end{aligned}$$

For further information about the Chebyshev polynomials and its recurrences you can check [1, Chapter 22], or [4, Chapter 18].

To compute the associated Legendre polynomials, we will use the following recurrences, which can be found in the cited references,

$$P_{\ell+1}^{\ell+1}(x) = -(2\ell - 1)\sqrt{1 - x^2}P_\ell^\ell(x), \tag{8}$$

$$P_{\ell+1}^{\ell}(x) = x(2\ell + 1)P_\ell^\ell(x), \tag{9}$$

$$(\ell - m + 1)P_{\ell+1}^m(x) = (2\ell + 1)xP_\ell^m(x) - (\ell + m)P_{\ell-1}^m(x). \tag{10}$$

More precisely we will use (8) to compute until P_m^m, then (9) to compute P_{m+1}^m, and finally (10) to compute until P_ℓ^m, using the $\cos\theta = z$ that we already know.

Now we have a precise way of calculating every term of the Fourier series, since we just need to repeat these calculations for every triangle of our triangulation and we will get the Fourier coefficient, and this way is how it has been implemented in our program. Finally to plot the surface we just apply (1),

$$\begin{aligned} S &= \{r^L(\varphi, \theta)\Phi(\varphi, \theta) \mid (\varphi, \theta) \in (0, 2\pi)\times(0, \pi)\}, \\ r^L(\varphi, \theta) &= \sum_{(\ell,m)\in L} r_\ell^m Y_\ell^m(\varphi, \theta). \end{aligned}$$

It is also interesting to see how we compute the formula for the estimated L^2-error given the triangulation. As seen in the formulas (6) and (7) all we need to compute are the norms squared. For the first norm we will use the following formula,

$$||r||^2 = \int_{\mathbb{S}^2} r^2 \, dV_g = \sum_{i=0}^{N} \int_{T_i^p} r^2 \, dV_g \approx \sum_{i=0}^{N} ||\overline{T}_i||^2 A(T_i^p).$$

For the norm of the differential we will use the scalar product as described giving us the following expression,

$$||dr||^2 = \int_{\mathbb{S}^2} g^*(dr, dr) \, dV_g = \int_{\mathbb{S}^2} \left[r_\theta^2 + \left(\frac{r_\varphi}{\sin\theta}\right)^2 \right] dV_g,$$

which can be approximated as shown in the thesis by the following sum,

$$||dr||^2 \approx \sum_{i=0}^{N} ||\overline{T}_i||^2 \left[\frac{||\overline{T}_i||^2}{(N \cdot \overline{T}_i)^2} - 1 \right] A(T_i^p).$$

We are just missing an expression for the Laplacian norm. In this case we will be using the formula discussed in [6], this will give us a value for the Laplacian in every vertex of the triangulation.

Let $\{v_i\}_{i=1}^V$ be the set of vertex of our triangulation, we denote $\{p_i\}_{i=1}^V$ their projections on the unit sphere so that $p_i = v_i^p$. For every vertex p_i with n neighbors, we denote $N(i) = i_1, \dots, i_n$ the set of index of the neighbor vertex of p_i. We will assume they are ordered counterclockwise as seen from outside the sphere above p_i. Then we can approximate the Laplacian in this spot as

$$\Delta_{\mathcal{M}} r(p_i) = \frac{4\sum_{j\in N(i)}(\cot\alpha_{ij} + \cot\beta_{ij})(||v_j|| - ||v_i||)}{\sum_{j\in N(i)}(\cot\alpha_{ij} + \cot\beta_{ij})||p_j - p_i||^2},$$

where α_{ij} and β_{ij} are the angles of the adjacent triangles to the segment p_ip_j.

Finally since we want to compute $||\Delta r||$. We will consider that each one of our vertex takes a region $s(p_i) \subset \mathbb{S}^2$ equivalent to one third of the area of the spherical triangles surrounding it. Therefore the area of all the triangles will be equally shared between the vertex, and we obtain the expression

$$\begin{aligned}||\Delta r||^2 &= \int_{S^2} (\Delta r)^2 \, dV_g \approx \sum_{i=0}^{V} \int_{s(p_i)} (\Delta_{\mathcal{M}} r)^2 \, dV_g \\ &= \sum_{i=0}^{V} \left[\frac{4\sum_{j\in N(i)}(\cot\alpha_{ij} + \cot\beta_{ij})(||v_j|| - ||v_i||)}{\sum_{j\in N(i)}(\cot\alpha_{ij} + \cot\beta_{ij})||p_j - p_i||^2} \right]^2 \frac{1}{3} \sum_{j\in N(i)} A(T_{ij}^p).\end{aligned}$$

8. Conclusion and results

As mentioned previously this thesis was built around the idea of creating a program that can replicate all the formulas explained here and can show some interesting results. For the interested reader this program does exist and can be found at `https://github.com/MiquelNasarre/FourierS2`.

It is satisfying to see that all these formulas actually work and can produce some interesting results. This can be seen in Figure 3, that shows the spherical harmonics as depicted by the program, and Figure 1, that shows a basic example of the program's functionality.

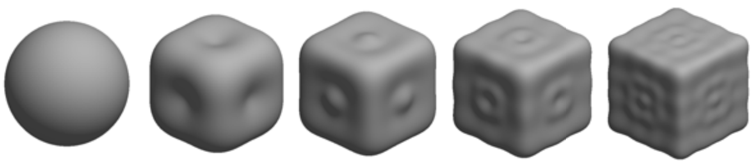

Figure 1: Program trying to recreate a cube triangulation with ℓ from 0 to 4.

Also this program allows us to see some clear limitations of the formulas. For example if you try to go too deep and your triangulation is not fine enough the approximations to calculate the coefficients will not be as good, giving you some weird looking shapes in the process, as seen in the middle shape of Figure 2. This limitation though can be easily solved by dividing the triangles in the triangulation, as shown by the last shape of Figure 2, where the shape is visibly better defined and the L^2-error is lower.

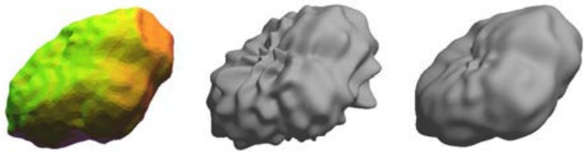

Figure 2: Program creating the Fourier series $\ell \leq 20$ of `example.dat` without the subdivision and with four subdivisions of the triangles.

Another limitation that can not be solved easily is the error formulas, due to the amount of approximations involving the entire process these formulas have proven not to be very reliable for the case in $L^2(\mathbb{S}^{\not{=}})$ although they have shown great results for the common Fourier series.

Overall I am very satisfied of the results obtained by the program as well as all the mathematical background developed in the thesis to back it up, I hope this article is useful for someone who decides to undertake a similar case of study in the future.

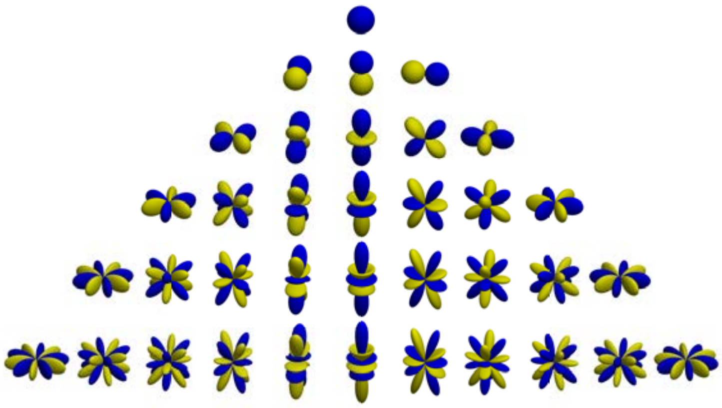

Figure 3: Spherical harmonics as shown by the program with ℓ from 0 to 5.

References

[1] M. Abramowitz, I.A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, For sale by the Superintendent of Documents, National Bureau of Standards Applied Mathematics Series **55**, U. S. Government Printing Office, Washington, DC, 1964.

[2] Y. Kosmann-Schwarzbach, *Groupes et symétries. Groupes finis, groupes et algèbres de Lie, représentations*, Second edition, Les Éditions de l'École Polytechnique, Palaiseau, 2006.

[3] D. Marín, M. Nicolau, A priori L^2-error estimates for approximations of functions on compact manifolds, *Mediterr. J. Math.* **12(1)** (2015), 51–62.

[4] F.W.J. Olver, D.W. Lozier, R.F. Boisvert, C.W. Clark (ed.), *NIST Handbook of Mathematical Functions*, U.S. Department of Commerce, National Institute of Standards and Technology, Washington, DC; Cambridge University Press, Cambridge, 2010. `dlmf.nist.gov`.

[5] S. Rosenberg, *The Laplacian on a Riemannian Manifold. An Introduction to Analysis on Manifolds*, London Math. Soc. Stud. Texts **31**, Cambridge University Press, Cambridge, 1997.

[6] G. Xu, Discrete Laplace–Beltrami operator on sphere and optimal spherical triangulations, *Internat. J. Comput. Geom. Appl.* **16(1)** (2006), 75–93.

REPORTS@SCM

AN ELECTRONIC JOURNAL OF THE
SOCIETAT CATALANA DE MATEMÀTIQUES

Extended Abstracts

Societat Catalana de Matemàtiques

Institut d'Estudis Catalans

SCM Master Thesis Day

Last October 3, with a notable attendance, we celebrated the third SCM TFM day. This is an activity organized by the Catalan Mathematical Society (SCM) which aims to facilitate those who have just graduated from a master's degree in mathematics at a Catalan university or from the common linguistic area (Xarxa Vives) to present their Final Master's Thesis. This interuniversitary activity it is about giving young master's graduates the opportunity to participate and present their first communication at a workshop, to energize the community of young mathematicians in the country that start the research, to inform about the convocation of the Galois awards and about the magazine *Reports@SCM*, and to spread the word about the mathematics master's programs of the universities of the Vives Network to students in the final year of the mathematics degree attending the day.

The day was held at the headquarters of the Institut d'Estudis Catalans and had the participation as speakers of eight students, and also with the presentation of the two master's theses awarded with the Evariste Galois 2025 prize (Pedro López and Joaquim Duran, winner and recipient), an award given by the SCM to the best final master's thesis of the previous year, in this case, 2024. In fact, one of the two winners of the 2025 Galois award presented their TFM in the 2024 edition of the SCM TFM day.

The scientific committee of the day was Enric Cosme (co-coordinator of the Master's in Mathematical Research of the UV-UPV), Simone Marchesi (editor in chief of *Reports@SCM*), Xavier Massaneda (coordinator of the Master's in Advanced Mathematics of the UB-UAB), Jordi Saludes (coordinator of the Master's in Advanced Mathematics of the UPC) and Pablo Sevilla (co-coordinator of the Master's in Mathematical Research of the UV-UPV). The organizing committee was Montserrat Alsina (president of the SCM), Josep Vives (vice-president of the SCM) and Òscar Burés and Philip Pita, current PhD students and former participants in previous editions of the day.

Reports@SCM collects in this issue the extended abstracts of the presentations of the day.

Extended Abstracts

REPORTS@SCM
AN ELECTRONIC JOURNAL OF THE
SOCIETAT CATALANA DE MATEMÀTIQUES

Densities for Hausdorff measure and rectifiability. Besicovitch's 1/2-conjecture

Jaume Capdevila Jové
Universitat Autònoma de Barcelona
jaume.capdevila.jove@uab.cat

Resum *(CAT)*
En aquest treball estudiem un dels conceptes centrals de la teoria geomètrica de la mesura, el de conjunt rectificable, i la seva relació amb les densitats per la mesura de Hausdorff. En aquesta interacció hi ha un dels problemes oberts més antics de la teoria: la conjectura-1/2 de Besicovitch. Estudiem una selecció de resultats rellevants, des dels articles pioners de Besicovitch [1] fins a la millora de Preiss i Tišer [7]. Després, presentem una contribució original: generalitzem a $\mathbb{R}^n$ un exemple donat originalment per Besicovitch en el pla, demostrant-ne les propietats clau i estenent així una cota inferior de la conjectura a dimensió arbitrària.

Keywords: *geometric measure theory, Hausdorff measure, rectifiability, Besicovitch's* 1/2*-conjecture.*

Abstract

One of the main concepts of geometric measure theory is that of m-rectifiable subsets of $\mathbb{R}^n$, given integers $0 < m \leq n$. They appear as a generalization of the notion of "nice" m-dimensional surfaces, such as C^1 submanifolds, or Lipschitz graphs. They are sets which, up to a set of zero $\mathcal{H}^m$-measure, are contained in a countable union of images of Lipschitz maps with domain in $\mathbb{R}^m$ (where $\mathcal{H}^m$ denotes the m-dimensional Hausdorff measure). For example, for $m = 1$, the 1-rectifiable sets are those which are contained in a countable union of rectifiable curves, again up to a set of zero $\mathcal{H}^1$-measure. On the other side of the coin, we have the purely m-unrectifiable sets, which are those that contain no m-rectifiable subset of positive $\mathcal{H}^m$-measure. One of the goals of geometric measure theory is to characterize rectifiability in terms of other geometric or analytical properties.

To that end, one of the basic tools is that of the densities for the Hausdorff measure. Consider a set $E \subset \mathbb{R}^n$ such that $0 < \mathcal{H}^s(E) < \infty$ for some $0 \leq s \leq n$, which we call an s-set. One defines the upper and lower s-densities of E at a point $x \in \mathbb{R}^n$, denoted as $\Theta^{*s}(E,x)$ and $\Theta^s_*(E,x)$ respectively, as the lim sup and lim inf as $r \to 0$ of

$$\frac{\mathcal{H}^s(E \cap B_r(x))}{(2r)^s}.$$

When both quantities coincide, the limit is called the s-density of E at x.

The densities for the Hausdorff measure and the notion of rectifiability are intimately connected. One of the most important theorems in this direction states that an m-set $E \subset \mathbb{R}^n$ is m-rectifiable if and only if the m-density of E exists and is equal to 1 at $\mathcal{H}^m$-almost all points of E. This is known as the characterization of rectifiability in terms of densities. This line of study was initiated in the pioneering work of Besicovitch [1] in 1938, where he established the result for 1-sets in the plane, i.e., the case $m = 1$ and $n = 2$. It was extended to arbitrary dimension in different stages, with the work of Moore [6], Marstrand [4] and Mattila [5].

Another point of connection between the two topics involves the lower density alone. It was proven by Besicovitch in the same article that if $\Theta^1_*(E, x) > 3/4$ for $\mathcal{H}^1$-almost all points of a 1-set E, then E is automatically 1-rectifiable. Following this idea, we define the following coefficient:

$$\sigma_m(\mathbb{R}^n) := \min\{\sigma > 0 : \text{for any } m\text{-set } E \subset \mathbb{R}^n,\ \Theta^m_*(E, x) > \sigma\, \mathcal{H}^m\text{-a.e. } x \in E \implies E \text{ is } m\text{-rectifiable}\}.$$

The previously stated result of Besicovitch translates to the bound $\sigma_1(\mathbb{R}^2) \leq 3/4$. Moreover, in the same article in 1938 he provided an example of a purely 1-unrectifiable set P which satisfies $\Theta^1_*(P, x) = 1/2$ at $\mathcal{H}^1$-almost all $x \in P$; a formal proof of this fact appeared later in a paper by Dickinson [3] in 1939. This way, they proved the lower bound $\sigma_1(\mathbb{R}^2) \geq \frac{1}{2}$. With this in mind, Besicovitch conjectured that the exact value of $\sigma_1(\mathbb{R}^2)$ is $1/2$, which is now known as *Besicovitch's* $1/2$*-conjecture*.

Further improvements to this bound have been obtained since then. In 1992, Preiss and Tišer [7] refined the estimate to $\sigma_1(\mathbb{R}^n) \leq (2 + \sqrt{46})/12 < 59/80$, which holds for all $n \geq 2$ (for all metric spaces, in fact). Recently, in 2024, Camillo De Lellis et al. [2] established that $\sigma_1(\mathbb{R}^n) \leq 7/10$, which is currently the best known upper bound.

In higher dimensions (for $m > 1$), no good upper bounds are known for $\sigma_m(\mathbb{R}^n)$. On the other hand, the same lower bound remains valid; in this work, we generalize Besicovitch's example in the plane to arbitrary dimensions, thereby showing

$$\sigma_m(\mathbb{R}^n) \geq \frac{1}{2}, \quad \text{for any } 0 < m < n.$$

This is an original contribution from this work.

References

[1] A.S. Besicovitch, On the fundamental geometrical properties of linearly measurable plane sets of points (II), *Math. Ann.* **115(1)** (1938), 296–329.

[2] C. De Lellis, F. Glaudo, A. Massaccesi, D. Vittone, Besicovitch's 1/2 problem and linear programming, Preprint (2024). `arXiv:2404.17536`.

[3] D.R. Dickinson, Study of extreme cases with respect to the densities of irregular linearly measurable plane sets of points, *Math. Ann.* **116(1)** (1939), 358–373.

[4] J.M. Marstrand, Hausdorff two-dimensional measure in 3-space, *Proc. London Math. Soc. (3)* **11** (1961), 91–108.

[5] P. Mattila, Hausdorff m regular and rectifiable sets in n-space, *Trans. Amer. Math. Soc.* **205** (1975), 263–274.

[6] E.F. Moore, Density ratios and $(\phi, 1)$ rectifiability in n-space, *Trans. Amer. Math. Soc.* **69** (1950), 324–334.

[7] D. Preiss, J. Tišer, On Besicovitch's $\frac{1}{2}$-problem, *J. London Math. Soc. (2)* **45(2)** (1992), 279–287.

REPORTS@SCM
AN ELECTRONIC JOURNAL OF THE
SOCIETAT CATALANA DE MATEMÀTIQUES

Fusion theorems and applications

Luis Pablo Colmenar
Universitat de València
cololuis@alumni.uv.es

Resum *(CAT)*
En teoria de grups finits, molts resultats clàssics impliquen subgrups de Sylow. Una direcció natural és generalitzar-los mitjançant subgrups de Hall. En aquest treball, mostrem com un resultat de Wielandt permet fer-ho eficaçment. Presentem dues aplicacions: una relacionada amb el teorema de fusió d'Alperin, i una altra amb el subnormalitzador, un concepte menys conegut però amb connexions recents amb la teoria de caràcters.

Societat Catalana de Matemàtiques
Institut d'Estudis Catalans

Keywords: *fusion in groups, subnormalizer.*

Abstract

In finite group theory, many results are formulated in terms of Sylow subgroups and rely heavily on the classical Sylow theorems. These results are central to the local-global philosophy of the subject, where local properties of subgroups provide valuable information about the structure of the whole group.

Whenever such theorems are established, a natural line of inquiry arises: can these results be generalized beyond Sylow subgroups? One promising direction involves replacing Sylow subgroups with Hall subgroups, which are more general but retain many desirable properties when they exist. However, such generalizations often require more sophisticated tools, since the theory of Hall subgroups is not as robust or widely applicable as Sylow theory in general finite groups.

In this work, we focus on a classical but perhaps underappreciated result by Wielandt, which proves to be a powerful instrument in extending certain Sylow-based statements to more general contexts involving Hall subgroups. Wielandt's theorem offers a unifying perspective that opens the door to new applications.

We present two main applications of this approach. The first concerns Alperin's fusion theorem, first proved by Alperin in [1], a fundamental result describing how conjugacy in a Sylow p-subgroup is controlled in terms of the local structure. This theorem is important in some conjectures in representation theory and character theory. We will show how Wielandt's result can be used to extend aspects of this theorem beyond the Sylow subgroups, providing a more flexible framework for studying fusion phenomena.

The second application involves a less well-known concept: the subnormalizer of a subgroup. This notion, mainly studied by Carlo Casolo in [2], tries to generalize the concept of normalizer. Subnormalizers offer an alternative lens through which one can examine the internal structure of a finite group. Recent developments show that this concept is not merely technical: it is connected to new conjectures in character theory and may lead to fresh insights into the interplay between subgroup structure and representation theory.

Both applications illustrate how classical tools, when viewed from a modern perspective, can be effectively repurposed to approach contemporary problems in group theory. The ideas we present highlight the ongoing relevance of results like Wielandt's theorem and demonstrate the value of re-examining classical results through new conceptual frameworks.

Acknowledgements

I would like to express my sincere gratitude to Alexander Moretó and Noelia Rizo for their valuable guidance, insightful comments, and careful corrections throughout the development of this work.

References

[1] J.L. Alperin, Sylow intersections and fusion, *J. Algebra* **6** (1967), 222–241.

[2] C. Casolo, Subnormalizers in finite groups, *Comm. Algebra* **18(11)** (1990), 3791–3818.

REPORTS @SCM
AN ELECTRONIC JOURNAL OF THE
SOCIETAT CATALANA DE MATEMÀTIQUES

Atypical values of complex polynomial functions

Manuel García García
Universitat de València
magarg25@alumni.uv.es

Resum *(CAT)*
Des de 1983, amb el treball de Broughton, s'han introduït diverses condicions de regularitat a l'infinit per a un polinomi complex f que garanteixen l'absència de valors crítics a l'infinit, és a dir, de valors atípics de f que no són valors crítics. En aquest treball recollim les condicions de regularitat més rellevants i estudiem les relacions que hi ha entre elles. En particular, responem a dues preguntes obertes proposades per Dũng Tráng Lê i J.J. Nuño-Ballesteros a [3].

Societat Catalana de Matemàtiques
Institut d'Estudis Catalans

Keywords: *complex polynomials, atypical values, critical values.*

Abstract

The topology of complex polynomial functions $f\colon \mathbb{C}^n \longrightarrow \mathbb{C}$ has been object of considerable study in recent decades. In particular, a central goal is to understand how the topology of the fibers $f^{-1}(c)$, $c \in \mathbb{C}$, changes. In this context, the concept of *locally trivial fibrations* plays a key role. Specifically, if f is locally a trivial fibration at $c \in \mathbb{C}$, then the topology of the fibers near c remains unchanged. The points $c \in \mathbb{C}$ where f fails to be locally a trivial fibration are called atypical values of f. The set of all atypical values of f is denoted by $\operatorname{Atyp} f$. In [4], Thom proved the finiteness of the set of atypical values. However, determining precisely this set is a major open problem.

Among the atypical values, one has the critical values, *i.e.*, $f(\Sigma f) \subset \operatorname{Atyp} f$, where Σf is the set of points $x \in \mathbb{C}^n$ where $\mathrm{d}f_x = 0$. In general, this inclusion is strict. Over the past decades, several *regularity conditions at infinity* for f have been introduced in order to guarantee the equality $f(\Sigma f) = \operatorname{Atyp} f$.

The first one is the notion of *tameness*, which was introduced by Broughton in [1] and [2]. In [5], Tibăr compiles some other regularity conditions at infinity, such as the *Malgrange Condition* (which generalizes the notion of tameness) and the *ρ-regularity at infinity*, where ρ is a *control function*. The following chain of implications is well-known:

$$\begin{aligned} f \text{ is tame} &\Longrightarrow f^{-1}(c) \text{ satisfies the Malgrange Condition} \\ &\Longrightarrow f^{-1}(c) \text{ is } \rho_E\text{-regular at infinity.} \end{aligned}$$

Most recently, in [3], Dũng Tráng Lê and J.J. Nuño Ballesteros introduced the notion of *atypical values from infinity*. In this paper, they generalize the Broughton's Global Bouquet Theorem in [2]. The paper concludes by posing several open questions aimed at gaining a deeper understanding of atypical values from infinity. Namely,

1. Is it true that, if f is tame, then f does not have atypical values from infinity?
2. Does a fiber $f^{-1}(c)$ which satisfies the Malgrange Condition correspond to a value c which is not an atypical value from infinity?

In this work we review all these definitions and explain our main contribution:

$$f^{-1}(c) \text{ is } \rho\text{-regular at infinity} \implies c \text{ is not an atypical value from infinity.}$$

Using this result, we obtain an extension of the previous chain of implications:

$$\begin{aligned} f \text{ is tame} &\implies f^{-1}(c) \text{ satisfies the Malgrange Condition} \\ &\implies f^{-1}(c) \text{ is } \rho_E\text{-regular at infinity} \\ &\implies c \text{ is not an atypical value from infinity.} \end{aligned}$$

This answers the first question of the authors in [3] and gives the right implication for the second one. The other implication remains open.

Acknowledgements

The author would like to deeply thank Professor Juan Antonio Moya Pérez for his encouragement during this Master's Thesis. The author would like to express his sincere gratitude to Professor J.J. Nuño-Ballesteros for his valuable advice and comments.

References

[1] S.A. Broughton, On the topology of polynomial hypersurfaces, in: *Singularities, Part 1* (Arcata, Calif., 1981), Proc. Sympos. Pure Math. **40**, American Mathematical Society, Providence, RI, 1983, pp. 167–178.

[2] S.A. Broughton, Milnor numbers and the topology of polynomial hypersurfaces, *Invent. Math.* **92(2)** (1988), 217–241.

[3] Dũng Tráng Lê, J.J. Nuño Ballesteros, A remark on the topology of complex polynomial functions, *Rev. R. Acad. Cienc. Exactas Fís. Nat. Ser. A Mat. RACSAM* **113(4)** (2019), 3977–3994.

[4] R. Thom, Ensembles et morphismes stratifiés, *Bull. Amer. Math. Soc.* **75** (1969), 240–284.

[5] M. Tibăr, *Polynomials and Vanishing Cycles*, Cambridge Tracts in Math. **170**, Cambridge University Press, Cambridge, 2007.

REPORTS@SCM
AN ELECTRONIC JOURNAL OF THE
SOCIETAT CATALANA DE MATEMÀTIQUES

Exploring the principles of coexistence in invader-driven replicator dynamics

Marina Garcia-Romero
Universitat Politècnica
de Catalunya (UPC)
marinagr99@gmail.com

Resum *(CAT)*
En aquest treball, utilitzem la "replicator equation" per explorar una de les qüestions fonamentals de la biologia evolutiva i l'ecologia: com es genera i es manté la biodiversitat? Centrant-nos en els sistemes "invader-driven", en què les interaccions o "fitnesses" de les espècies estan determinades per l'espècie invasora independentment de l'espècie envaïda, busquem relacionar les "fitnesses" amb les espècies que coexisteixen als estats finals d'equilibri. Descobrim el mecanisme que regeix la selecció d'espècies supervivents i que maximitza la resistència del sistema envers les invasions externes, i trobem que el nombre mitjà d'espècies que coexisteixen creix amb el nombre inicial d'espècies.

Societat Catalana de Matemàtiques
Institut d'Estudis Catalans

Keywords: *biological modelling, replicator equation, pairwise invasion fitness matrix, multi-species system, coexistence, invasion resistance.*

Abstract

Studying the non-linear and often complex dynamics of large systems of interacting species, competing or cooperating between them, can help to discover the principles that, in ecosystems, lead some species to survive and coexist, while others go extinct, to better understand of one of the central questions in ecology and evolutionary biology that remains unsolved, which is how biodiversity is generated and maintained. In the early 1970s, ecologists widely accepted that the stability and resilience observed in rich ecosystems were enhanced by complexity and biodiversity, until in 1972 the paradigm shifted completely when Robert May mathematically showed that random complexity tends to destabilise system dynamics [3]. This raised a contradiction between observation and theory known as the ecology paradox or diversity-stability debate, highlighting the need for some hidden structure or pattern in nature, such as the antisymmetric prey-predator interactions [1]. In this work, we study invader-driven interactions as a potential mechanism for the stabilization of large complex systems and we find that, under certain assumptions, invader-driven systems lead to the coexistence of species.

We use the replicator equation as a theoretical framework [4], which originated in game theory but has been widely applied to biology and epidemiology [2]. Given a system with N species, $S = \{1, 2, \dots, N\}$, consider the pairwise invasion fitness λ_i^j from species i to j, with $i, j \in S$ and $\lambda_i^i = 0$, and the invasion fitness matrix $\Lambda = (\lambda_i^j)_{i,j\in S}$. Then, the replicator equation models the time evolution of the species frequencies $\mathbf{z} = (z_1, z_2, \dots, z_N)$, with $\sum_{i\in S} z_i = 1$ and $0 \leq z_i \leq 1\ \forall\, i \in S$, as

$$\dot{z}_i = \Theta z_i((\Lambda\mathbf{z})_i - \mathbf{z}^\top\Lambda\mathbf{z}) = \Theta z_i\left(\sum_{j\neq i}\lambda_i^j z_j - Q(\mathbf{z})\right), \quad i \in S,$$

where the constant $\Theta \geq 0$ is the speed of dynamics and $Q(\mathbf{z})$ is the global mean fitness or system resistance to external invasion. In particular, we focus on invader-driven systems, in which pairwise interactions are determined by the invading species regardless of the invaded one, i.e., $\lambda_i^j = \lambda_i(1 - \delta_i^j)\ \forall\, i, j \in S$, so each species i is characterized by its active trait λ_i. We study the equilibrium states $\mathbf{z}^*$ to understand how fitnesses in the case $\lambda_i > 0\ \forall\, i \in S$ relate to the set of surviving or coexisting species at equilibrium, $S^* = \{i \in S \mid z_i^* > 0\} \subseteq S$, finding that $Q^* = Q(\mathbf{z}^*)$ plays a crucial role in the species selection process.

We prove that locally asymptotically stable equilibria are always composed of the top $n = |S^*|$ species without gaps, with fitnesses $\lambda_N \leq \cdots \leq \lambda_n \leq \cdots \leq \lambda_2 \leq \lambda_1$ and $2 \leq n \leq N$, and, furthermore, we find numerical evidence that each system contains just one of these equilibria, which in turn is a global attractor (any initial condition with all species present tends asymptotically towards it). Therefore, for each S there is a unique S^* that can be asymptotically reached by the dynamics and, hence, a unique set of species characterized by n that end up coexisting. We discover the mechanism ruling the species selection (see Figure 1), which starting with two species iteratively adds a species $i \in S$ if $Q_{i-1}^* < \lambda_i$, that is, if it can invade the previous $i - 1$, until some species n meets the condition $\lambda_{n+1} < Q_n^* < \lambda_n$. Moreover, we prove that in each step Q^* increases, $Q_{i-1}^* < Q_i^*$, so this biological process tends to maximize the system resistance to invasion.

Lastly, using this mechanism we create an algorithm that allows to find n for several invader-driven systems generated randomly with $\lambda_i \sim \mathcal{U}[0, 1]$, from $N = 5$ to $N = 500$. Fitting the data we find that the mean number of coexisting species increases according to $\bar{n} = 1.381\sqrt{N}$, suggesting that invader-driven interactions could be a potential mechanism through which ecosystems stabilize and maintain biodiversity.

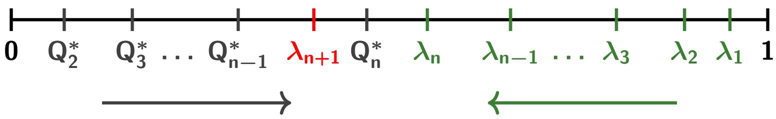

Figure 1: Species selection mechanism in invader-driven systems, $0 < \lambda_N \leq \cdots \leq \lambda_n \leq \cdots \leq \lambda_2 \leq \lambda_1 \leq 1$.

Acknowledgements

I want to thank Erida Gjini for welcoming me into her group and for all the interesting discussions, and Sten Madec, Nicola Cinardi, Tomás Freire, Tomás Camolas and Thao Le, for creating an enriching environment. I also want to thank José Tomás Lázaro Ochoa for his co-supervision, support and important contributions.

References

[1] T. Chawanya, K. Tokita, Large-dimensional replicator equations with antisymmetric random interactions, *J. Phys. Soc. Japan* **71(2)** (2002), 429–431.

[2] S. Madec, E. Gjini, Predicting N-strain coexistence from co-colonization interactions: epidemiology meets ecology and the replicator equation, *Bull. Math. Biol.* **82(11)** (2020), Paper no. 142, 26 pp.

[3] R.M. May, Will a large complex system be stable?, *Nature* **238(5364)** (1972), 413–414.

[4] P.D. Taylor, L.B. Jonker, Evolutionarily stable strategies and game dynamics, *Math. Biosci.* **40(1-2)** (1978), 145–156.

REPORTS@SCM
AN ELECTRONIC JOURNAL OF THE
SOCIETAT CATALANA DE MATEMÀTIQUES

Unique preduals and free objects in Banach spaces

Mario Guillén
Universitat Politècnica de València (UPV)
mguisan3@alumni.upv.es

Pedro Tradacete
Instituto de Ciencias Matemáticas (ICMAT)

Societat Catalana de Matemàtiques
Institut d'Estudis Catalans

Resum *(CAT)*
Estudiem quan un espai de Banach té un únic predual, centrant-nos primer en les funcions holomorfes acotades al disc unitat i analitzant la demostració d'Ando. Considerem com estendre el resultat a diverses variables, on apareixen dificultats tècniques. També tractem diferents condicions suficients per garantir la unicitat i el cas de reticles de Banach.

Keywords: *unique predual, Banach space, bounded holomorphic functions, Property (X), L-embedded space, free Banach lattice.*

Abstract

A *predual* of a Banach space X is a Banach space Y such that there exists an isomorphism $Y^* \to X$. When X admits only one such space Y up to isometric isomorphism, we say that X has a *unique predual*. The problem of determining when a Banach space has a unique predual is a central one in functional analysis, starting in the works of Dixmier (1948) and Ng [5] and studied by Sakai, Ando, Godefroy, Pfitzner, and others. Classical examples of spaces with unique preduals include von Neumann algebras by Sakai (1971), the space of bounded holomorphic functions on the complex disk $H^\infty(\mathbb{D})$ by Ando [1], and separable L-embedded Banach spaces by Pfitzner [6]. Godefroy's survey [3] remains a key reference summarizing these developments and listing open problems.

This project revisits the uniqueness problem with emphasis on the space of bounded holomorphic functions, $H^\infty(U)$, defined on an open subset $U \subset \mathbb{C}^n$. We review Ando's original proof of the uniqueness of the predual of $H^\infty(\mathbb{D})$, which identifies the space $L^1(\mathbb{T})/H^1_0(\mathbb{T})$ as its unique predual. We present a detailed proof following both Ando's original formulation [1] and a later one from Garnett [2], filling several gaps left unproved in the literature.

We extend Ando's result to the case where U is a disjoint union of simply connected open subsets of the complex plane. Using Mujica's notion of the holomorphic free Banach space $G^\infty(U)$ given in [4], characterized by the universal property

$$H^\infty(U, F) \cong L(G^\infty(U), F),$$

we explicitly construct an isometric isomorphism

$$G^\infty(U) \cong \bigoplus_{\alpha \in A} G^\infty(U_\alpha),$$

where $U = \bigsqcup_{\alpha\in A} U_\alpha$. This allows us to prove that $H^\infty(U)$ has a unique predual whenever U is such a disjoint union, thus generalizing Ando's result.

We then attempt to extend the result to several complex variables, considering $H^\infty(\mathbb{T}^n)$ and $H^\infty(B_n)$. Following a different path of proving uniqueness in the one-dimensional case, we reduce the problem of proving that $H^\infty(B^n)$ has a unique predual to showing that $B_{H_0^1(\mathbb{S}_n)}$ is $\|\cdot\|_p$-closed for some $p \in (0,1)$. However, this higher-dimensional setting presents several challenges. When passing from one variable to several, the space $H_0^1(\mathbb{S}_n)$ is no longer contained in the Hardy space $H^1(\mathbb{B}_n)$, so pre-compactness arguments from Hardy space theory cannot be applied. Consequently, no conclusive results were obtained in this direction.

This work also reviews techniques guaranteeing uniqueness of preduals in broader settings. Two sufficient conditions are revisited: *Property (X)* (Godefroy–Talagrand, 1980), which ensures that X is the unique predual of X^*, and the notion of *L-embedded spaces*, for which separable cases were solved by Pfitzner [6]. Finally, we explore the Banach lattice setting, introducing the *free Banach lattice* FBL$[E]$ (Avilés–Rodríguez–Tradacete, 2015). Although we explore possible definitions for predual equivalence in this setting, we find difficulties, particularly because the space of lattice homomorphisms is not a vector space, thus leaving this problem for future research.

Acknowledgements

This project has been developed under the JAE Intro Scholarship from the Consejo Superior de Investigaciones Científicas (CSIC).

References

[1] T. Ando, On the predual of H^∞, *Comment. Math. Spec. Issue* **1** (1978), 33–40.

[2] J.B. Garnett, *Bounded Analytic Functions*, Revised first edition, Grad. Texts in Math. **236**, Springer, New York, 2007.

[3] G. Godefroy, Existence and uniqueness of isometric preduals: a survey, in: *Banach Space Theory* (Iowa City, IA, 1987), Contemp. Math. **85**, American Mathematical Society, Providence, RI, 1989, pp. 131–193.

[4] J. Mujica, Linearization of bounded holomorphic mappings on Banach spaces, *Trans. Amer. Math. Soc.* **324(2)** (1991), 867–887.

[5] K.F. Ng, On a theorem of Dixmier, *Math. Scand.* **29** (1971), 279–280.

[6] H. Pfitzner, Separable L-embedded Banach spaces are unique preduals, *Bull. Lond. Math. Soc.* **39(6)** (2007), 1039–1044.

Extended Abstracts

REPORTS@SCM
AN ELECTRONIC JOURNAL OF THE
SOCIETAT CATALANA DE MATEMÀTIQUES

Idempotent elements of the group algebra

Vicent Miralles Lluch
Universitat Politècnica de València
vicent@vmiralles.com

Societat Catalana de Matemàtiques
Institut d'Estudis Catalans

Resum *(CAT)*
L'objectiu d'aquest treball és estudiar els elements idempotents centralment primitius de l'àlgebra de grup i desenvolupar un mètode per al seu càlcul en el cas de cossos finits. A partir de la teoria de representacions de grups finits i de resultats sobre mòduls, àlgebres i extensions de cossos, s'introdueix el concepte de cos d'escissió per a un grup. Finalment, s'explora com l'acció de Galois sobre l'àlgebra de grup definida sobre aquests cossos permet obtindre aquests idempotents del cos original.

Keywords: *idempotent elements, splitting fields, group algebra, Galois action, finite fields.*

Abstract

This work focuses on the study of idempotent elements of group algebras, with particular emphasis on centrally primitive idempotents. These elements are fundamental because they allow the algebra to be decomposed into simpler blocks. The importance of centrally primitive idempotents lies in the fact that each of them generates one of these blocks and, moreover, they form a basis for the centre of the algebra, which completely defines its structure.

The main objective is to develop an explicit and practical method for calculating these idempotents over fields whose characteristic does not divide the order of the group (which we will assume to be finite), and which are often not algebraically closed. This is no easy task, since many results in representation theory rely on the latter property (see [3]) and are not valid in a more general context. For this reason, we resort to the concept of a splitting field for a group (see [1, 2]), which generalises the algebraically closed field, providing a theoretical framework that guarantees the validity of many classical results, including the expression of these idempotents.

The method we develop, often known as Galois descent, consists of exploiting the expression of centrally primitive idempotents of the group algebra over a splitting field. The idea is to consider a finite Galois extension of the original field that is a splitting field for the group; in this extension, the Galois group acts on these idempotents. The expression of these idempotents is known since they are defined over a splitting field, and it can be shown that the sum of the orbits resulting from this action ultimately gives us the centrally primitive idempotents we are looking for in the original field (see [4]). This method is significantly simpler than other approaches, such as the one in [5], which relies on the computation of a division ring's dimension—a generally non-trivial task. Furthermore, we prove that both methods are equivalent by summing over a general orbit to obtain the expression given in [5].

To illustrate the procedure, we conclude with a detailed application to finite fields, where the efficiency of our approach becomes particularly evident. The practicality of the method lies not only in the simplicity of the orbit computations—thanks to the cyclic nature of the Galois group generated by the Frobenius automorphism—but also in the theoretical results previously developed in this work, which directly provide the corresponding splitting fields. In this example, we first establish the identification between characters over the splitting field and ordinary characters via Brauer characters (see again [3]), and then carry out the explicit computation of the idempotents, thereby demonstrating the applicability and strength of our self-contained approach.

Acknowledgements

I would like to express my gratitude to María José Felipe Román, Víctor Manuel Ortiz Sotomayor and Xaro Soler Escrivà for allowing me to develop this work, as well as for their guidance and corrections.

References

[1] C.W. Curtis, I. Reiner, *Representation Theory of Finite Groups and Associative Algebras*, Pure Appl. Math. **XI**, Interscience Publishers (a division of John Wiley & Sons, Inc.), New York-London, 1962.

[2] K. Doerk, T. Hawkes, *Finite Soluble Groups*, De Gruyter Exp. Math. **4**, Walter de Gruyter & Co., Berlin, 1992.

[3] I.M. Isaacs, *Character Theory of Finite Groups*, Pure Appl. Math. **69**, Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London, 1976.

[4] G. Karpilovsky, *Group Representations. Volume 1. Part B: Introduction to Group Representations and Characters*, North-Holland Math. Stud. **175**, North-Holland Publishing Co., Amsterdam, 1992.

[5] K. Lux, H. Pahlings, *Representations of Groups. A Computational Approach*, Cambridge Stud. Adv. Math. **124**, Cambridge University Press, Cambridge, 2010.

Extended Abstracts

REPORTS@SCM
AN ELECTRONIC JOURNAL OF THE
SOCIETAT CATALANA DE MATEMÀTIQUES

On nilpotency in braces and the Yang–Baxter equation

Alberto Rodríguez Durà
Universitat de València
albertord.bg@gmail.com

Resum *(CAT)*
Les brides són estructures algebraiques que permeten estudiar les solucions no degenerades de l'equació de Yang–Baxter (EYB). Cada brida admet una solució no degenerada i, recíprocament, tota solució d'aquest tipus està determinada per una brida associada. Així, la classificació de les solucions no degenerades depén de l'anàlisi estructural de les brides. Les seues propietats algebraiques es corresponen amb les de les solucions, i la nilpotència permet descriure el caràcter multipermutacional d'aquestes estructures.

Societat Catalana de Matemàtiques
Institut d'Estudis Catalans

Keywords: *braces, Yang–Baxter, nilpotency.*

Abstract

The Yang–Baxter equation (YBE) is a fundamental equation in theoretical physics, arising independently in the works of C. N. Yang (1967), Nobel Laureate in Physics, and R. J. Baxter, within the frameworks of quantum field theory and the study of integrable models in statistical mechanics, respectively.

The formulation of the YBE is strongly inspired by the celebrated Reidemeister moves (cf. [3]).

Consequently, the study of YBE solutions has gained significant relevance in recent decades, both because of its intrinsic importance and its applications in braid theory, braided groups, quantum groups, cryptography, and noncommutative geometry.

The multidisciplinary context of the YBE has generated great interest in the search for and classification of its solutions.

Open Problem. To find and classify the solutions of the Yang–Baxter equation.

Given the Herculean nature of this task, the Fields Medalist V. G. Drinfeld ([2]) proposed focusing on the so-called set-theoretic solutions of the YBE, a type of combinatorial solution whose geometric and symmetric character naturally gives rise to algebraic techniques.

In this work, we undertake a thorough analysis of the algebraic property of nilpotency in braces, as a clear and significant example of the translation of algebraic properties into classificatory properties of YBE solutions. We study the so-called lateral nilpotencies in braces, which have a distinct impact both on the structural analysis of braces and on the classification of solutions. In this context, a key concept of nilpotency in braces—one that has recently emerged and has a decisive impact both structurally within braces and classificatorily within YBE solutions—is central nilpotency in braces. This type of nilpotency arises with the aim of unifying both lateral nilpotencies in braces, as shown in [1], where it is demonstrated that central nilpotency in braces can be regarded as the true analogue, within brace theory, of group nilpotency.

Within group theory, the local study of nilpotency or p-nilpotency associated with a prime p has undergone substantial development following the seminal works of Hall and Higman (cf. [4]). A key concept in this context is the p-Fitting subgroup of a finite group, the largest normal p-nilpotent subgroup of the group.

The main objective and contribution of this work is the introduction and analysis of central p-nilpotency in finite braces. We conduct a comprehensive structural study of central p-nilpotency in braces, allowing us to define an appropriate p-Fitting ideal. This contribution is original within the theory and is intended to inspire further developments in this field.

Acknowledgements

The author gratefully acknowledges Professor Adolfo Ballesteros and Professor Vicent Pérez (Universitat de València) for insightful guidance and constructive feedback throughout this research.

References

[1] A. Ballester-Bolinches, R. Esteban-Romero, M. Ferrara, V. Pérez-Calabuig, M. Trombetti, Central nilpotency of left skew braces and solutions of the Yang–Baxter equation, *Pacific J. Math.* **335(1)** (2025), 1–32.

[2] V.G. Drinfeld, On some unsolved problems in quantum group theory, in: *Quantum Groups* (Leningrad, 1990), Lecture Notes in Math. **1510** Springer-Verlag, Berlin, 1992, pp. 1–8.

[3] P. Etingof, T. Schedler, A. Soloviev, Set-theoretical solutions to the quantum Yang–Baxter equation, *Duke Math. J.* **100(2)** (1999), 169–209.

[4] P. Hall, G. Higman, On the p-length of p-soluble groups and reduction theorems for Burnside's problem, *Proc. London Math. Soc. (3)* **6** (1956), 1–42.

REPORTS@SCM
AN ELECTRONIC JOURNAL OF THE
SOCIETAT CATALANA DE MATEMÀTIQUES

Density of hyperbolicity in families of complex rational maps

Francesc Timoner Vaquer
Universitat de Barcelona
francesc.timoner-vaquer@utoulouse.fr

Societat Catalana de Matemàtiques
Institut d'Estudis Catalans

Resum *(CAT)*
Un dels problemes oberts centrals és la densitat d'hiperbolicitat. En aquest treball ho investiguem en la dinàmica complexa unidimensional, i ens concentrem en el cas polinòmic (cas particular d'una funció racional) com a model on els mecanismes principals poden ser exposats i comprovats en detall. La via procedimental és clara: primer, la construcció de peces de puzle en un entorn del conjunt de Julia; segon, l'ús d'aquestes per definir una funció de caixa complexa; i finalment, l'aplicació de teoremes de rigidesa a aquestes. Aquest procés tradueix la informació combinatòria en rigidesa per als polinomis, demostrant que un polinomi no renormalitzable pot ser aproximat per un polinomi hiperbòlic.

Keywords: *complex dynamics, holomorphic dynamics, rational maps, hyperbolicity, renormalisation, complex box mapping.*

Abstract

A *rational map* is a holomorphic analytic function $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ on the Riemann sphere that can be written as the quotient of two coprime polynomials; equivalently, $f(z) = \frac{P(z)}{Q(z)}$, where P, Q are complex polynomials of some degree. The degree of f is defined as $d = \max(\deg P, \deg Q)$, and we assume that $d \geq 2$. In the particular case where Q is a constant, f is just a polynomial. Rational maps of degree $d \geq 2$ form a finite-dimensional space, so exploring this parameter space is feasible. Every rational map of degree $d \geq 2$ has $2d-2$ critical points (counting multiplicity), and near these points, the map behaves like $z \mapsto z^k$, so it is highly contracting and fails to be injective. Away from the critical points, f is a local homeomorphism.

Definition (Hyperbolic rational map)**.** A rational map is said to be *hyperbolic* if all its critical points are in the basins of attracting periodic points.

Conjecture (Density of hyperbolicity)**.** *The hyperbolic rational maps form an open and dense set in the space of all rational maps of a given degree.*

Definition. We say that a map is *non-renormalisable* if it does not admit any polynomial-like restriction for any iteration with connected filled-in Julia set.

The main goal of this thesis is to deconstruct, understand all details and prove the following theorem:

Main Theorem (Theorem 1.3 in [3])**.** *Let f be a non-renormalisable polynomial of degree $d \geq 2$, without neutral periodic points. Then, f can be approximated by a sequence of hyperbolic polynomials (g_i) of the same degree.*

To tackle the problem we review and combine several fundamental tools: puzzle piece decompositions so we can consider returns and track symbolically critical orbits, Böttcher coordinates near infinity that linearize escaping behaviour, holomorphic motions to follow dynamical objects across parameters, and quasi-conformal conjugacies to transfer geometric control between maps. Also, we suppose our map is non-renormalisable: for a rational map, one demands that its critical orbits do not return in small neighbourhoods in a "periodic way". These maps are often rigid, in the sense that their combinatorial structure determines their geometry. Finally, and most importantly, we make use of complex box mappings as an induced map defined on a disjoint union of topological discs that captures return dynamics of critical orbits inside a controllable domain ("upgrade" of the famous polynomial-like maps). These are flexible enough to encode both local renormalisation behaviour and global combinatorial constraints.

The seminal paper [3] (our main reference) lacks explicit technical assumptions in its statements and proofs. This makes some of their statements, as written in that paper, not entirely correct. Some assumptions were implicit or not considered, for example, the dynamically natural property of complex box mappings. Some parts, claimed to be straightforward, are not. In [1] they clarified and fixed some results on rigidity of polynomials and box mappings, but the theorem stated above remains unclear. So for the first time in the literature of complex dynamics, we provide detailed explanations for each part of the proof of that theorem. We consider the implications and ensure validity, especially when considering the dynamically natural property of box mappings. Our aim is to review the existing literature ([4, 2]), emphasising crucial aspects, and comprehensively understand the tools required for the theorem's proof. We aim to encapsulate them in a "black-box" and use them to advance research, for instance, to establish the density of hyperbolicity in other families of rational maps. We believe this meticulous deconstruction and attention to detail can significantly contribute to the general public's understanding of the subject matter.

The proof of the Main Theorem lies on a construction of dynamically natural box mappings for non-renormalisable polynomials without neutral periodic points together with a verification of the hypotheses needed to invoke rigidity theorems. In rigid families, topologically conjugate maps are automatically more regular (e.g., quasi-conformal or conformal in complex dynamics). By another of the main theorems needed (Theorem 6.1 in [1]), combinatorially equivalent non-renormalisable dynamically natural complex box mappings are rigid, and hence, quasi-conformally conjugate. This result, along with other known or basic notions, leads to the quasi-conformal rigidity of non-renormalisable polynomials. Consequently, the original polynomial is approximated, in the uniform topology on compact sets, by hyperbolic polynomials; hence density of hyperbolicity holds for the class considered.

References

[1] T. Clark, K. Drach, O. Kozlovski, S. van Strien, The dynamics of complex box mappings, *Arnold Math. J.* **8(2)** (2022), 319–410.

[2] K. Drach, D. Schleicher, Rigidity of Newton dynamics, *Adv. Math.* **408** (2022), part A, Paper no. 108591, 64 pp.

[3] O. Kozlovski, S. van Strien, Local connectivity and quasi-conformal rigidity of non-renormalizable polynomials, *Proc. Lond. Math. Soc. (3)* **99(2)** (2009), 275–296.

[4] C.T. McMullen, D.P. Sullivan, Quasiconformal homeomorphisms and dynamics. III. The Teichmüller space of a holomorphic dynamical system, *Adv. Math.* **135(2)** (1998), 351–395.

Reports@SCM
Volume 10, number 1, 2025

Table of Contents