Università di Padova – Dipartimento di Matematica "Tullio Levi-Civita"

Scuole di Dottorato in Matematica Pura e Computazionale

Seminario Dottorato 2025/26

Preface

This document offers an overview of the activity of Seminario Dottorato 2025/26.

Our "Seminario Dottorato" (Graduate Seminar) has a double purpose. At one hand, the speakers — usually Ph.D. students or post-docs, but sometimes also senior researchers — are invited to communicate their researches to a public of mathematically well-educated but not specialist people, by preserving both understandability and the flavour of a research report. At the same time, people in the audience enjoy a rare opportunity to get an accessible but also precise idea of what's going on in some mathematical research area that they might not know very well.

Let us take this opportunity to warmly thank once again all the speakers for having held these interesting seminars and for their nice agreement to write down these notes to leave a concrete footstep of their participation.

We are also grateful to the collegues who helped us, through their advices and suggestions, in building an interesting and culturally complete program.

Padova, June 20th, 2026

Corrado Marastoni, Tiziano Vargiolu

Abstracts (from Seminario Dottorato's webpage)

Thursday 6 November 2025

Letting AI untie the Knots

Ludovico Morellato (Padova, Dip. Mat.)

Links are one-dimensional smooth manifolds embedded in $\mathcal{S}^3$, and they are central objects in low-dimensional topology; for instance, every closed, connected, oriented 3-manifold can be obtained via Dehn surgery along a link in $\mathcal{S}^3$.
The main goal of this report is to show how Reinforcement Learning (RL) can be used to obtain new upper bounds on the unknotting number $u(\cdot)$ and the slice genus $g_4(\cdot)$ of knots and links, two topological invariants for which no general algorithm is known. We first introduce the relevant background in Knot Theory, then sketch the RL framework, and finally present the bounds we obtained for several families of knots and links.
From a joint work with Dai, Hayman and Juhász.

Thursday 20 November 2025

Isoperimetric problem in the Heisenberg group

Davide Massenz (Padova, Dip. Mat.)

The isoperimetric problem asks which sets minimize perimeter among all sets with a fixed volume. This is one of the most ancient questions in mathematics, dating back to the legend of Queen Dido. Despite the ancient origins, a turning point came with the work of De Giorgi only around the 1950s. He introduced a general definition of perimeter in the n-dimensional Euclidean space and this allowed for a more rigorous formulation of the isoperimetric problem. Later the problem has been generalized by other mathematicians to different frameworks, like Riemannian manifolds and metric spaces.
In this talk, after an historical introduction, we will at first recall the well-known isoperimetric inequality in the euclidean case, where the solution is the ball, examining in detail the notion of perimeter on which it is based. We will then introduce the Heisenberg group as the simplest and most studied example of a sub-Riemannian manifold and we will discuss the isoperimetric problem in this setting, highlighting Pansu's conjecture which provides the candidate isoperimetric sets, known as Pansu spheres.

Wednesday 17 December 2025

Evolving geometries: an introduction to Mean Curvature Flow

Gaia Bombardieri (Padova, Dip. Mat.)

How do shapes evolve when driven by their mean curvature? In Euclidean space, Huisken's classical theorem ensures that convex surfaces become spherical as they shrink to a point.
In this talk, we investigate the behavior of the flow in the sub-Riemannian setting of the Heisenberg group H^1. We will first introduce the basics of classical Mean Curvature Flow, focusing on the role of self-shrinkers as models for singularities. Then, we will move to the sub-Riemannian context, analyzing the flow for mean convex hypersurfaces. We will discuss the problem of selfsimilarity in this setting and present a recent result: the Pansu sphere, despite being the candidate isoperimetric profile of H^1, is not a self-shrinker. This reveals a striking divergence from the Euclidean intuition, where the static isoperimetric solution and the dynamic evolution profile coincide.

Thursday 29 January 2026

Effect of bond disorder on the supercritical discrete Gaussian free field

Emanuele Pasqui (Padova, Dip. Mat.)

The discrete Gaussian free field is a model of random variables on graphs that is central in statistical mechanics, and can be interpreted as a microscopic description of fluctuations in a homogeneous crystal at non-zero temperature.
In this introductory talk, we define the model when the dimension is larger than 2, and we discuss the so-called hard wall event, in which the field is constrained to stay non-negative on a given set. We study both the decay of its probability and the behaviour of the field conditioned on this event, as the size of the set grows to infinity. Finally, we introduce disorder into the model in the form of random weights on the edges of the graph, modelling inhomogeneities in the crystal, and analyse how this disorder affects the aforementioned results.
This talk is based on a joint work with Alberto Chiarini.

Thursday 12 February 2026

Motivic Character sheaves as the "center" of the universal Hecke category

Runlei Xiao (Padova, Dip. Mat.)

Beginning with Harish-Chandra, a recurring theme in the representation theory of reductive groups is that characters behave like central elements: they act by scalars and naturally live in the "middle" of convolution algebras. The geometric developments of the 1970s and 80s — most notably Deligne-Lusztig theory and Lusztig's introduction of character sheaves — shifted this idea from functions to sheaves and, in doing so, turned centrality into a built-in feature of the formalism. Convolution is promoted to a monoidal structure on equivariant derived categories, and the horocycle (Harish-Chandra) transform is realized geometrically as a pull-push functor along explicit correspondences. In this talk I will unfold this story in the concrete case of $SL(2)$. I'll work through hands-on calculations that make the geometric argument visible, aiming to isolate the key intuition behind "characters as central objects" in a way that remains accessible throughout. Meanwhile, i will

explain how motive involves in this story.

Thursday 26 February 2026

Profinite groups and probability

Nowras Naufel (Padova, Dip. Mat.)

Profinite groups are a class of topological groups which show up in a broad range of mathematical subjects and,as such, they can be studied with tools coming from algebra, analysis, geometry, probability, among others.
Most of the talk will be spent introducing the concept of profinite groups and highlighting key examples and properties. Time permitting, I will briefly speak about how probability can come into play.

Thursday 12 March 2026

In Search of Tilting Representations

Anastasios Slaftsos (Padova, Dip. Mat.)

The representation theory of quivers (i.e., oriented graphs encoding algebraic data) consists a central pillar in the study of finite-dimensional algebras. Beyond their combinatorial nature, quivers provide concrete models for more abstract settings, including module categories and their derived counterparts. A fundamental problem in this area is the classification and comparison of representations of a fixed quiver, and tilting theory provides a key mechanism for approaching this problem. In this talk, we introduce tilting theory through explicit examples, beginning with classical tilting modules over finite dimensional algebras and progressing to more general triangulated settings that extend beyond the classical derived category framework.

Thursday 26 March 2026

An introduction to infinite-dimensional geometry in continuum mechanics

Francesca Berlinghieri (Padova, Dip. Mat.)

Continuum mechanics studies the motion and deformation of bodies modeled as continuous media. It covers a large number of theories, including ideal, compressible and viscous fluid models, as well as linear and nonlinear elasticity.
In this talk, a transition from Newtonian mechanics of discrete systems to the setting of continuum mechanics will be presented through the principle of virtual works. Particular emphasis will be given to the role of the placement (or deformation) map as primary descriptor of the continuum. Such maps naturally belong to functional spaces that are inherently infinite-dimensional. This perspective motivates the study of the geometry of spaces of admissible deformations. As a classical example, we will briefly discuss the approach introduced by Arnold, and later developed by Ebin and Marsden, where the motion of an incompressible perfect fluid is interpreted as a geodesic flow

on the (infinite dimensional) Lie group of volume-preserving diffeomorphisms.

Thursday 9 April 2026

Isoperimetric Inequalities Between Two Fractional Perimeters

Giacomo Cozzi (Padova, Dip. Mat.)

In this talk I will introduce the notion of fractional perimeter of a set, show its main properties and motivate the interest in this object. Then, I will discuss an ongoing project in which we prove that the ball is a local minimizer of the scale invariant isoperimetric ratio between two fractional perimeters under small graph perturbations: this parallels the results known in the literature where one of the two fractional perimeters is replaced by the volume.
This is a joint work with Annalisa Massaccesi (University of Padova), Giovanni Alberti and Jeremy Mirmina (University of Pisa).

Thursday 23 April 2026

Methods for Complex Network Analysis

Matteo Bergamaschi (Padova, Dip. Mat.)

The spread of information across complex networks has emerged as a central topic in today's data-driven world, giving rise to numerous challenging optimization problems.
In this talk, I will introduce two fundamental problems in this domain: influence maximization and technology diffusion. Both problems focus on identifying key agents (nodes) that are most effective at propagating influence throughout a network. After a short review of classical approaches used to address these problems, I will introduce a novel methodology, Network Adaptive Direct Search. Finally, I will discuss a potential extension of the influence maximization framework to applications in public finance.

Thursday 7 May 2026

A Fast Ride on the Turnpike

Nicolò De Bernardi (Padova, Dip. Mat.)

The term "turnpike" first appeared in econometrics in the 1950s, although the phenomenon had been observed by Von Neumann in 1938. It describes a stability property of optimal trajectories: the optimal path between any two points is the one that moves rapidly from the initial condition toward a steady state (the "turnpike"), remains close to it for most of the time horizon, and adjusts to meet the terminal condition only near the final time, provided the endpoints are sufficiently far apart. This property has been established in various contexts, including control theory, dynamical systems and Mean Field Games, under some suitable monotonicity assumptions on the model.
In this talk we will present some basic examples of turnpike properties, with the aim of highlighting the main structural features of the models that ensure their validity. Moreover, in the context of

quadratic MFG, we will introduce a weaker monotonicity assumption that still guarantees the turnpike behavior toward local equilibria, along with an interpretation in terms of spectral properties of an operator encoded in the MFG system.

Thursday 4 June 2026

An Introduction to Scattered Data Histopolation

Giacomo Cappellazzo (Padova, Dip. Mat.)

Kernel-based methods provide a powerful and flexible framework for addressing histopolation problems, where the input data consists of average values of a function over intervals or higher-dimensional regions rather than pointwise samples.
This presentation introduces the concept of Averaging Kernel Hilbert Spaces (AKHS), a natural extension of classical Reproducing Kernel Hilbert Spaces (RKHS) tailored to histopolation. Within this framework, we develop construction principles for averaging kernels. We establish a fundamental isometric isomorphism between an AKHS and an associated RKHS, which allows reformulating histopolation as an interpolation problem in the corresponding RKHS. This connection yields explicit expressions for the histopolation matrix, criteria for unisolvence, and convergence results for the mean values of the histopolants. The presentation will briefly cover fundamental topics such as polynomial interpolation, Radial Basis Function (RBF) interpolation and the related challenges of optimizing node placement and the shape of basis functions.

Thursday 18 June 2026

Counting curves using partial differential equations

David Klompenhouwer (Padova, Dip. Mat.)

Enumerative geometry is the art of counting geometric objects that satisfy certain geometric conditions. For example: how many plane curves, given by polynomials of degree d, pass through a fixed set of 3d-1 distinct points in the plane? These kinds of questions have been asked since ancient times. Partial differential equations (PDEs), on the other hand, are a much more recent tool and are used to model all sorts of physical phenomena. In the early 1990s, physicist Edward Witten made a striking conjecture that revealed an intimate connection between these two research areas. Since then, the love between enumerative geometry and PDEs has blossomed, giving rise to fascinating new insights in both areas.
In this seminar, I will give a friendly introduction to both of these topics, by focusing on objects and ideas that are of particular interest to me and showing how they are related. At the end, I will mention how my research attempts to explain this mysterious romance between enumerative geometry and PDEs.

Letting AI untie the Knots

Ludovico Morellato (*)

Abstract. Links are one-dimensional smooth manifolds embedded in $\mathcal{S}^3$, and they are central objects in low-dimensional topology; for instance, every closed, connected, oriented 3-manifold can be obtained via Dehn surgery along a link in $\mathcal{S}^3$.
The main goal of this report is to show how Reinforcement Learning (RL) can be used to obtain new upper bounds on the unknotting number $u(\cdot)$ and the slice genus $g_4(\cdot)$ of knots and links, two topological invariants for which no general algorithm is known. We first introduce the relevant background in Knot Theory, then sketch the RL framework, and finally present the bounds we obtained for several families of knots and links.
From a joint work with Dai, Hayman and Juhász.

1 Knot Theory basics

1.1 Definitions

We can think of knots as "closed strings" in the three dimensions, as if we were to tie a knot with a shoelace in the usual way and then glue the two extremities of the string in order to get a closed loop. The two simplest examples are the *unknot* and the *trefoil knot*, depicted in Figure 1: some more intricate knots are collected in Figure 2, while a few "multi-component knots", called *links*, are shown in Figure 3.

Let us introduce these objects formally. The ambient space in which we work is the 3-dimensional sphere, $\mathcal{S}^3 = \left\{x \in \mathbb{R}^4 \mid \|x\| = 1\right\} \subseteq \mathbb{R}^4$. We can also think about it as the 3-dimensional space to which we add a point at infinity, $\mathcal{S}^3 = \mathbb{R}^3 \sqcup \{\infty\}$.

Intuitively, knots and links are pieces of string drawn in $\mathbb{R}^3$, but it is often convenient to work in the 3-sphere $\mathcal{S}^3$. Since knots are compact, the two choices give exactly the same theory (the equivalence classes are in canonical bijection); $\mathcal{S}^3$ is preferred because it is compact and has no distinguished point, which makes several constructions cleaner. In this report we will use $\mathcal{S}^3$ and $\mathbb{R}^3$ interchangeably.

(*)Ph.D. course, Università di Padova, Dip. Matematica, via Trieste 63, I-35121 Padova, Italy. E-mail: morellat@math.unipd.it. Seminar held on 6 November 2025.

Figure 1: The two simplest knots: the unknot (left) and the trefoil (right).

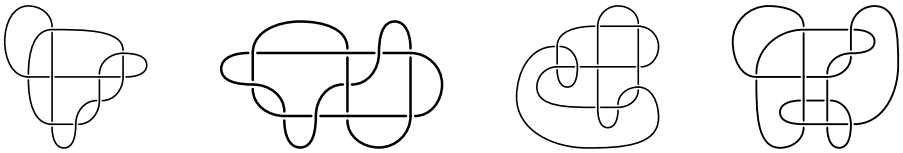


Figure 2: More complicated knots: 9_{17}, 10_{18}, 10_{65} and 11_{121}. The labels follow the standard Rolfsen notation (more details in Section 1.3)

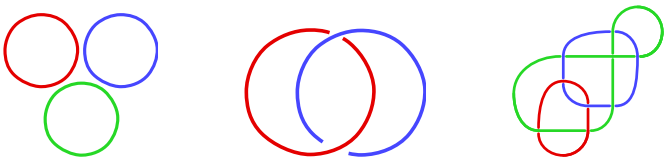


Figure 3: Three links: the 3-component unlink (left), the Hopf link $\texttt{L2a}_1$ (center) and $\texttt{L9a}_{50}$ (right).

Definition 1.1 (k-component link) A smooth *k-component link* in $\mathcal{S}^3$ is a smooth embedding

$$L : \bigcup_{i=1}^{k} \mathcal{S}^1 \times \{i\} \hookrightarrow \mathcal{S}^3.$$

For the most part, we are going to study knots, 1-component links.

Definition 1.2 (Knot) A smooth *knot* in $\mathcal{S}^3$ is a smooth embedding

$$K : \mathcal{S}^1 \hookrightarrow \mathcal{S}^3.$$

The image $\mathrm{Im}(L) \subset \mathcal{S}^3$ is a smooth one-dimensional submanifold of $\mathcal{S}^3$, to which we add the orientation that makes the map L orientation-preserving. Remark that it is also possible to retrieve L starting from $\mathrm{Im}(L)$: given an oriented smooth one-dimensional submanifold of $\mathcal{S}^3$, the map L is uniquely determined up to isotopy. Hence, we will usually identify L with its image, and we think of links as oriented smooth one-submanifolds of $\mathcal{S}^3$.

We usually denote the various connected components of L by L_i, i.e.

$$L = L_1 \sqcup \ldots \sqcup L_k.$$

1.2 Isotopies and link diagrams

We regard links up to *ambient isotopies*.

Definition 1.3 (Ambient isotopy) Let $L, L' \colon \bigcup_{i=1}^{k} \mathcal{S}^1 \hookrightarrow \mathcal{S}^3$ be two k-component links. An *ambient isotopy* between L and L' is a smooth map

$$\Phi \colon \mathcal{S}^3 \times [0,1] \longrightarrow \mathcal{S}^3$$

such that, writing $\Phi_t(x) := \Phi(x,t)$ for every $t \in [0,1]$,

- the map $\Phi_t \colon \mathcal{S}^3 \to \mathcal{S}^3$ is a diffeomorphism for every $t \in [0,1]$;
- $\Phi_0 = \mathrm{id}_{\mathcal{S}^3}$, i.e. the isotopy starts at the identity;
- $\Phi_1 \circ L = L'$, i.e. at time $t = 1$ the diffeomorphism Φ_1 carries L onto L'.

If such a Φ exists, we say that L and L' are *ambient isotopic* (or simply *equivalent*), and we write $L \sim L'$.

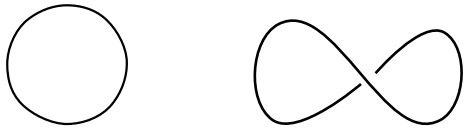


Figure 4: Two equivalent knots: the unknot and a "twisted" unknot.

Although links live in $\mathcal{S}^3$, they can be faithfully encoded by 2-dimensional pictures, called *link diagrams*. These are none other than what we used so far to represent links on paper, and we can rigorously define them.

Definition 1.4 (Diagram of a link) A *diagram* D of a link $L \subset \mathbb{R}^3$ is the projection of $\mathrm{Im}(L)$ onto a plane $\pi \subset \mathbb{R}^3$ such that:

- the projection has only finitely many multiple points;
- every multiple point is a transverse double point, called a *crossing* (no triple points, no tangencies);
- at every crossing, we record which of the two strands lies above the other in $\mathbb{R}^3$: the *over-strand* is drawn unbroken, while the *under-strand* is drawn with a small gap.

The over/under decoration at the crossings is exactly what is needed to recover L from D up to ambient isotopy. Indeed, starting from $D \subset \mathbb{R}^2$, lift each strand back to $\mathbb{R}^3$ by setting its third coordinate to 0 everywhere, except in a small neighbourhood of every crossing, where the over-strand is bumped slightly above the plane and the under-strand slightly below; the resulting curve in $\mathbb{R}^3$ is ambient isotopic to L, and any two such lifts (for different choices of bump heights) are ambient isotopic to each other. In other words, a diagram encodes the link up to equivalence.

The following result allows us to work at the level of diagrams in a more combinatorial-friendly way.

Theorem 1.5 (Reidemeister, [Cro04, Theorem 3.8.1]) *Two diagrams represent the same link if and only if one can be transformed into the other via a sequence of Reidemeister moves.*

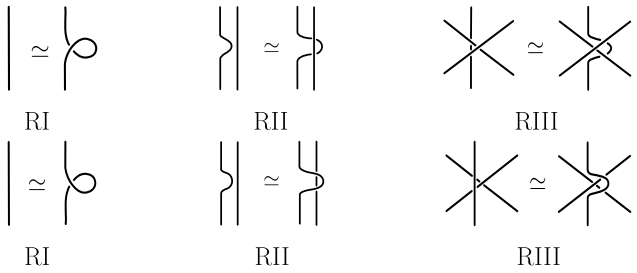


Figure 5: The three Reidemeister moves in the "under-strand" and "over-strand" versions.

Thanks to the above discussion, from this point onward we can freely speak about link diagrams in the plane instead of links in 3-dimensions without loss of generality; moreover, we consider diagrams up to Reidemeister moves.

Talking about diagrams is particularly convenient since it allows us to define a computer-friendly way to encode a knot diagram into an object called *PD code*.

Definition 1.6 (PD code of a diagram)
Let D be an oriented link diagram with n crossings. Label each crossing with an integer from 1 to n; we call them $C1, \dots, Cn$. Then start from a crossing and walk through the link following the orientation; while doing so, label each segment between two crossings with integers starting from 1 in increasing order. Repeat for every component, always increasing the segment labels.
At the end of this labelling procedure, we are going to have all the segments with numbers from 1 to $2n$.

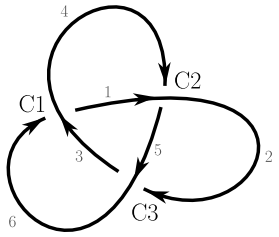


Figure 6: PD code of the trefoil: $[(6,3,1,4),(4,1,5,2),(2,5,3,6)]$

The *Planar Diagram code (PD code)* is the ordered list of n 4-tuples, where: the i-th tuple corresponds to the crossing Ci and it reports the labels of the four segments around the crossing, starting from the incoming under-strand and going around counter-clockwise.

Remark 1.7 Given a PD code of a link, it determines uniquely the link diagram. For multi-component links with few crossings on some components, it might happen that the orientation is not uniquely determined. This problem can be avoided by adding the data of crossing signs (± 1) to the PD code.

1.2.1 The isotopy problem

Understanding whether two diagrams represent the same link or not can be very difficult and is a central problem in Knot Theory. In Section 1.4 we will see how topological invariants come to help us in doing this. The following two results give an idea of what "difficult" means.

Theorem 1.8 (Coward and Lackenby, 2014, [CL14, Theorem 1.1]) *Given two diagrams of the same knot, let n be the sum of the numbers of crossings of the two diagrams. Then, one needs at most*

$$2^{2^{2^{\cdot^{\cdot^{n}}}}}$$

Reidemeister moves to deform one diagram into the other, where the height of the tower of 2s (with a single n at the top) is $10^{1{,}000{,}000 \cdot n}$.

Theorem 1.9 (Lackenby, 2015, [Lac15, Theorem 1.1]) *If a diagram of the unknot has c crossings, then there is a sequence of at most $(236 \cdot c)^{11}$ Reidemeister moves taking it to the trivial diagram. Moreover, every diagram in this sequence has at most $(7c)^2$ crossings.*

In practice, sometimes it is very easy to understand if two diagrams represent the same link, like in Figure 4, and sometimes it can be very difficult: see for instance Figure 7 for two very different and quite complicated representations of the unknot.

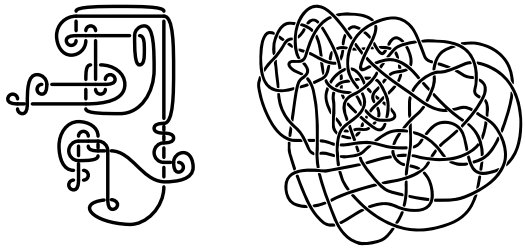


Figure 7: Two diagrams for the unknot.

1.3 Algebraic structure and tabulation

Knots have some nice algebraic structure: they can be summed via *connected sum*. Take two knots K_1 and K_2. Intuitively, to construct the connected sum $K_1 \sharp K_2$ means to take one string and tie first K_1, then K_2 and finally to glue the two ends together. See Figure 8 for an example.

Definition 1.10 (Connected sum of knots) Let K_1 and K_2 be two oriented knots in $\mathcal{S}^3$. For each K_i, choose a small 3-ball $B_i^3 \subset \mathcal{S}^3$ that meets K_i in a single unknotted arc. Cut that arc out of each knot, then glue the two complementary 3-balls along their boundary spheres in such a way that the four free endpoints of the cut arcs are matched in pairs, compatibly with the orientations of K_1 and K_2. The result is a single oriented knot in $\mathcal{S}^3$, the *connected sum* $K_1 \sharp K_2$. This operation is well-defined up to equivalence of knots.

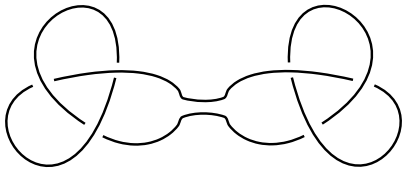


Figure 8: An example: connected sum of two trefoils (also called Square Knot).

A notion of *prime knot* follows naturally.

Definition 1.11 (Prime knot) A knot that cannot be decomposed as a connected sum of two non-trivial knots is said to be *prime*.

Furthermore, one can show that knots form a commutative monoid with unique factorisation. We can classify knots via the minimal number of crossings they can have in a diagram. There are tables classifying all prime knots up to some size. For instance, the (prime) knots up to 5 crossings are displayed in Table 1.

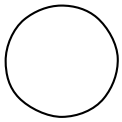
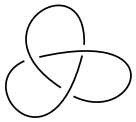
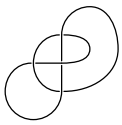
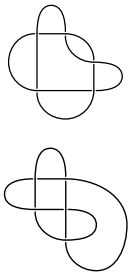
0 crossings	3 crossings	4 crossings	5 crossings
			
Unknot	Trefoil (3_1)	Figure Eight (4_1)	Cinquefoil (5_1), 3-twist knot (5_2)

Table 1: Rolfsen table for knots up to 5 crossings.

As the number of crossings increases, the number of prime knots grows at least exponentially; we can see the amount of prime knots up to 12 crossings in Table 2.

# crossings	0	3	4	5	6	7	8	9	10	11	12
# prime knots	1	1	1	2	3	7	21	49	165	552	2,176

Table 2: Rolfsen table for knots up to 12 crossings.

1.4 Link invariants

One goal of Knot Theory is to find a smart way to distinguish different links. This can be done with the use of *link invariants*, intrinsic topological properties of the link which do

not change when it is deformed.

Link invariants come in several flavours. *Numerical* invariants assign a number to each link. Some examples include the unknotting number $u(K)$ and the slice genus $g_4(K)$ that we will study in detail, but also the signature $\sigma(K)$, the Rasmussen s-invariant, the Ozsváth–Szabó τ-invariant, the Seifert genus $g_3(K)$, and many others. *Polynomial* invariants assign a Laurent polynomial to each link, the most classical examples being the Alexander and Jones polynomials. Finally, *homological* invariants such as Heegaard Floer and Khovanov Homology associate a graded chain complex to each link, from which several of the numerical invariants above are extracted. We will not define these invariants here, but they will reappear later as bounds for $u(K)$ and $g_4(K)$.

A simple example we can describe explicitly is **knot 3-colourability**. A knot is 3-colourable if it exists one diagram of it in which one can colour its strands with three colours so that at every crossing the strands are either all the same colour or all different; lastly, at least two colours must be used.

One can show that this is a topological invariant (it is sufficient to prove that this property is not compromised by Reidemeister moves). Therefore, if we have one knot that is 3-colourable and another that is not, we can be sure that they are distinct. In Figure 9 we can see that the Trefoil is 3-colourable and the Figure Eight is not, and hence they are not the same knot.

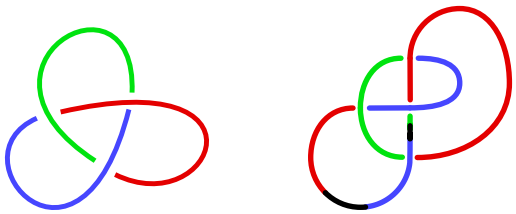


Figure 9: Tricolouring shows that Trefoil and Figure Eight are distinct knots.

1.5 The unknotting number

The unknotting number is a numerical invariant that wants to understand the beknottedness of a knot.

Definition 1.12 (Unknotting number) Let $K \subset \mathcal{S}^3$ be a knot. The *unknotting number of* K, denoted $u(K)$, is the minimum number of times that K must pass through itself to become the unknot.

For the sake of simplicity, let us start the discussion from diagrams. What we do is take a diagram D and we try to untie it by doing the simplest local change that impacts its beknottedness.

Definition 1.13 (Crossing change on a diagram) Let D be a diagram of a knot K and let c be a crossing of D. The *crossing change* at c is the local move that swaps which of the two strands at c lies above the other, leaving the rest of D unchanged.

And we define the unknotting number of a diagram to be the following.

Definition 1.14 (Unknotting number of a diagram) Let D be a diagram of a knot K. The *unknotting number of* D, $u(D)$ is the minimum number of crossing changes one has to make in order to obtain the unknot starting from D.

For example, the trefoil can be turned into the unknot by performing a single crossing change at one of its crossings (Figure 10); already at this stage we can read off that the trefoil can be untied by at most one such move. Hence, the unknotting number of the diagram displayed in Figure 10 is *at most 1*.

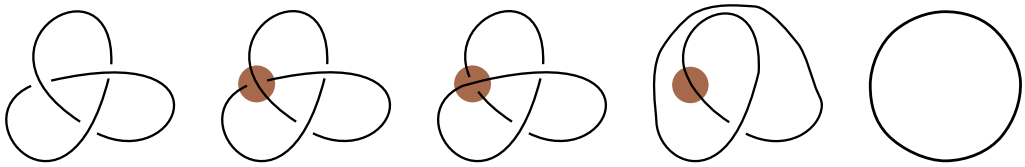


Figure 10: One crossing change unties the trefoil: starting from the diagram on the left, we perform a crossing change on the highlighted crossing; then, via Reidemeister moves and planar isotopies, we show that the diagram we obtained is the one of the unknot.

The number obtained on a single fixed diagram is, however, only an upper bound of the true unknotting number: changing the diagram we are working on, we could obtain a diagram for which the unknotting number is lower. Indeed, it holds for a knot K

$$u(K) = \min_{\substack{D \text{ diagram} \\ \text{for } K}} u(D),$$

i.e. computing the true unknotting number means to consider *all* diagrams of K, which are infinitely many.

To remove this dependence on a specific diagram while still working with combinatorial moves in dimension 2, we package the data of "deforming first, then perform a crossing change" into a single elementary object, called *band path*.

Definition 1.15 (Band path) A *band path* on a diagram of K is the data of:

- a portion of the diagram (a strand between two crossings of K);
- a sequence of Reidemeister II moves we want to perform on that strand, possibly interleaved with Reidemeister I moves; this is the "path" we are sliding the chosen strand on to reach a target strand of the diagram;

At the end of the band path, we are going to perform the crossing change at one of the two new crossings created by the last Reidemeister II move.

An example can be seen in Figure 11.

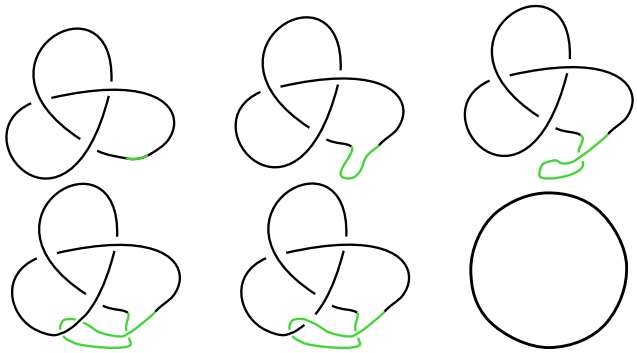


Figure 11: The trefoil has unknotting number at most 1: movie with a band.

A single crossing change on the original diagram can itself be represented by a band path: it is a single Reidemeister II move (with possibly one Reidemeister I move) done before the crossing change takes place. For instance, the band path and consecutive crossing change depicted in Figure 11 are equivalent to performing a single crossing change on the lower crossing of the diagram.

Band paths allow us to encapsulate the 3-dimensionality of the problem, free of the rigidity of a chosen diagram, in a 2-dimensional framework. Therefore, from now on when we say "we perform a crossing change", we actually mean that we are going to perform the instruction encapsulated into a band path.

It should also be clear from the above definitions that the unknotting number $u(K)$ of a knot K is the minimum number of band paths needed to turn (a diagram of) K into a diagram of the unknot.

Any diagram of K can be reduced to a diagram of the unknot by finitely many crossing changes (by choosing a starting point and switching crossings to obtain an ascending diagram, which is necessarily a diagram of the unknot; see, e.g., [Ada04, Discussion in Section 3.1]). For a non-trivial knot K we have $u(K) \geq 1$ (otherwise it would need to be trivial from the start), and the trefoil example above shows $u(K) \leq 1$, so $u(K) = 1$.

The computation of the unknotting number is not straightforward at all. For instance, out of the 249 knots with up to 10 crossings, we do not know the unknotting number of the following 10 (for all of them, we know it to be either 2 or 3): 10_6, 10_{11}, 10_{47}, 10_{51}, 10_{54}, 10_{61}, 10_{76}, 10_{77}, 10_{79}, 10_{100}.

1.6 The slice genus

A central theme in low-dimensional topology is the study of surfaces bounded by a given knot. In 1934, Herbert Seifert [Sei35] presented a constructive algorithm to produce such a surface $S \subset \mathbb{R}^3$ such that $\partial S = K$. The question he was trying to answer was: what is the minimum genus of a surface S such that $\partial S = K$? The answer is invariant under ambient isotopies and it is called the *(Seifert) genus of* K, denoted by $g_3(K)$.

Let us take this idea a step further. Let $B^4 = \{x \in \mathbb{R}^4 \mid \|x\| \leq 1\}$, so that $\mathcal{S}^3 = \partial B^4$, and fix a knot K in $\mathcal{S}^3$. So we can think of a knot as staying "on the border of the 4-dimensional ball" and we can try to construct a surface F in the interior of B^4 such that

$\partial F = K \subset \mathcal{S}^3$.

Definition 1.16 (Slice surface) A *slice surface* for a knot K is a compact, connected, oriented surface $F \subset B^4$, such that $\partial F = K$.

As in the 3-dimensional case, we ask ourselves what is the minimal genus such a surface can have. The answer is again a topological invariant, hence we define it formally.

Definition 1.17 (Slice genus) The *slice genus* $g_4(K)$ of K is the minimal genus of a slice surface for K.

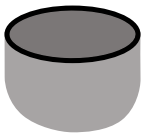


Figure 12: The slice genus (and the Seifert genus) of the unknot is zero: we can embed a disc D^2 in B^4 gluing its border on the unknot and obtaining a slice surface of genus zero for the unknot.

Remark that $g_4(K) \leq g_3(K)$: a Seifert surface can be embedded in B^4 and become a slice surface, however inside B^4 there is "more room" to deform than inside $\mathcal{S}^3$, the minimum can only decrease.

A useful way to visualize the construction of a slice surface is to think of it as a *movie* that unties our knot in four dimensions: indeed, in dimension four, every knot unties. We interpret the fourth coordinate of B^4 as time: at each instant $t \in [0, 1]$ we see a knot diagram living in a copy of $\mathbb{R}^2$, and as t flows forward we are allowed to deform that diagram by ambient isotopies in $\mathbb{R}^4$: planar isotopies, Reidemeister moves and non-trivial local change, called *band addition*, possible because of the fourth dimension.

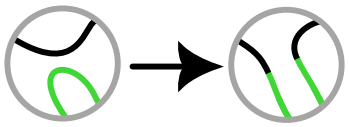


Figure 13: The band addition: the two adjacent strands are "cut and reconnected the other way". This move results in a saddle point on the surface, as we will see in Example 18.

The film starts at $t = 0$ with a diagram of our knot K and ends at $t = 1$ with a diagram of the unknot; then we glue a flat disc to the final unknot, identifying the border of the disc to the unknot, closing the bottom of the surface. The whole trace swept out by the moving diagram, together with this final cap, is a compact oriented surface $F \subset B^4$ with $\partial F = K$: a slice surface for K. Its genus is determined entirely by the moves we performed during the movie, so every such film gives an explicit upper bound on $g_4(K)$.

Even though at first sight the slice genus seemed to live in a fundamentally different world from the unknotting number, it should be now clear that the band framework that we developed for $u(K)$ in Section 1.5 can be reused almost verbatim to produce diagrammatic

upper bounds on $g_4(K)$ as well. The key is to perform band additions instead of crossing changes at the end of the band: instead of swapping the over/under information at the new crossing produced by the final Reidemeister II move of a band, we identify a 'target strand', we cut the target and the band and we reconnect them in a way that agrees with the orientation of the two strands (see Figure 14).

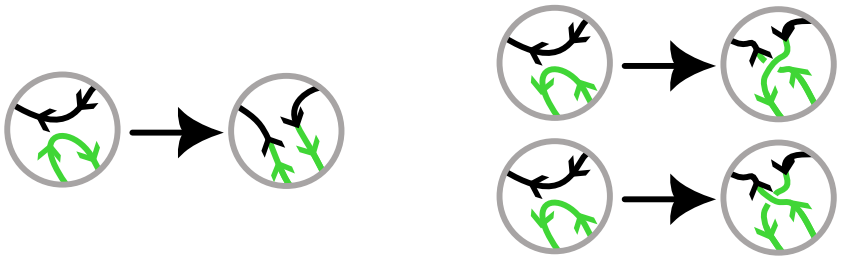


Figure 14: Consider the top string (the black one) to be the target strand and the bottom one (the green one) to be the final part of the band before the band addition takes place. The band addition has to agree with the orientations: on the left, no twist is needed, on the right we have two possible cases (we can think of them as performing a final Reidemeister I move before the band addition takes place).

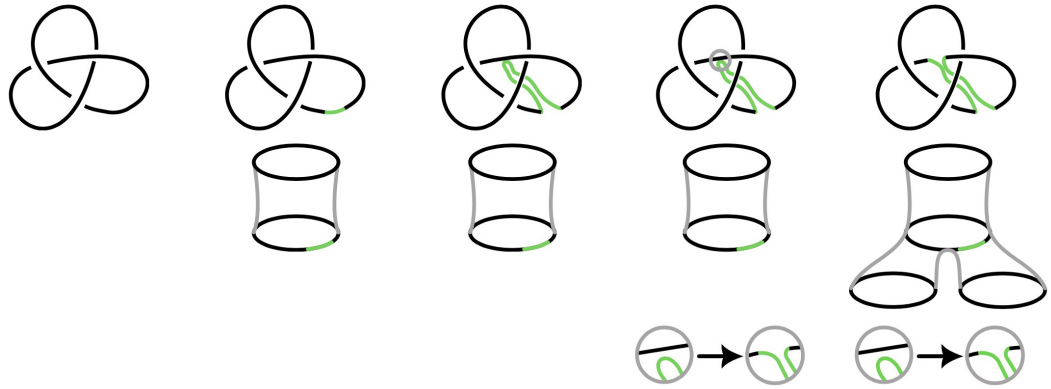
We can therefore play exactly the same game as before on a fixed diagram of K (choose a strand to deform, slide it next to a target strand via Reidemeister II move, possibly interleaved with Reidemeister I moves) but now we conclude each band with a band addition rather than a crossing change. A sequence of such band paths taking a diagram of K to the unknot produces a slice surface for K whose genus is an explicit upper bound on $g_4(K)$.

We regard the two computations as two variants of the same combinatorial game on a fixed diagram of K: choose a sequence of band paths (each ending either in a crossing change or in a band addition) that simplifies the diagram all the way to the unknot, and try to minimise the upper bound you obtain. This is exactly the setup we will feed to the RL agent in the next section.

Example 1.18 (Slice surface of genus 1 for the trefoil) As a concrete example, we study the case of the trefoil. Figures 15 and 16 show a *slice movie* for the trefoil. We are going to represent the surface F we are constructing and above it the present state of the knot while we modify it: this corresponds to the boundary component of the surface we are building that is not the original knot (i.e., the 'bottom boundary'). This is a link, which changes as time flows from 0 to 1; let us call it L_{t_x} for $t_x \in [0,1]$. While we represent L_{t_x} with a true diagram, we simplify the shape of the surface by representing the "horizontal sections" as circles.

Let us start with the trefoil $K = L_0$ depicted in Figure 15a; recall that K lives on $\mathcal{S}^3$, i.e. the boundary of B^4. We push the trefoil from $\mathcal{S}^3$ inside the ball B^4, obtaining at time t_b (Figure 15b) a cylinder of sort with base the trefoil. We select a portion of K on L_{t_b}, we highlight it in green. We continue pushing down the link, while deforming it with a Reidemeister II move until we reach the state depicted in Figure 15c. At this point, we can perform the *band addition* shown in Figure 15d; notice that this corresponds to adding

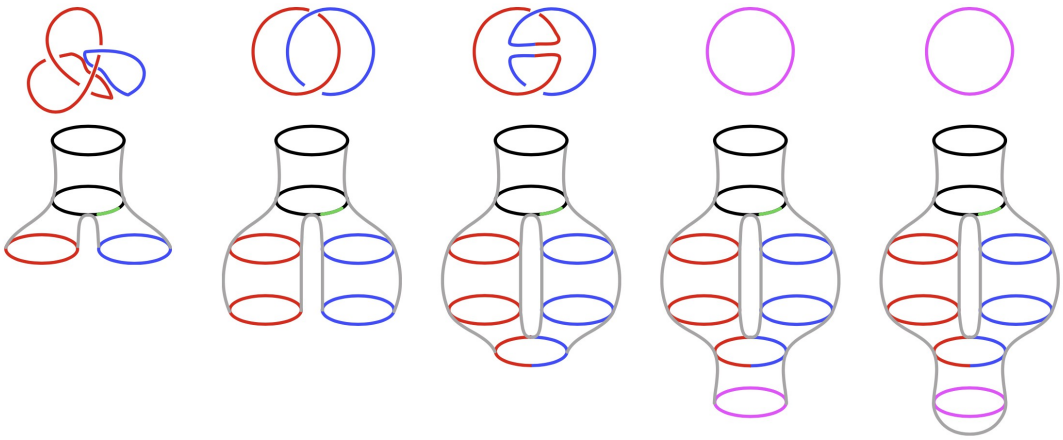
a saddle to F (Figure 15e).



(a) Time $t_a = 0$ (b) Time $t_b > t_a$ (c) Time $t_c > t_b$ (d) Time $t_d > t_c$ (e) Time $t_e > t_d$

Figure 15: Building a slice surface for the trefoil: first part.

As the surface now has two disconnected boundary components "on the bottom", the link L_{t_e} we obtained is a 2-component link. We colour the two different components in red and blue, as shown in Figure 16f. While we let the time flow until t_g, we perform some Reidemeister moves and we simplify the state of L_{t_f}, until we get L_{t_g} which is the Hopf link (Figure 16g). We can now perform another band addition on the link, which results in another saddle (Figure 16h). Remark that this saddle fuses the two different components, which we now colour in violet, and we can simplify L_{t_h} via Reidemeister moves obtaining L_{t_i}, the unknot (Figure 16i). At this point, we can cap-off the final boundary component by gluing a disc to it, obtaining a surface F embedded in B^4 for which $\partial F = K$; this can be seen in Figure 16j.



(f) Time $t_f > t_e$ (g) Time $t_g > t_f$ (h) Time $t_h > t_g$ (i) Time $t_i > t_h$ (j) Time $t_j = 1$

Figure 16: Building a slice surface for the trefoil: second part.

This shows that the slice genus of the trefoil is at most one, concluding our example.

The computation of the slice genus is, just as for the unknotting number, far from straightforward. For instance, out of the 9,988 prime knots with 13 crossings, we do not know the slice genus of 19 of them; an excerpt of this list is the following: $\texttt{13a}_{2554}$, $\texttt{13a}_{2844}$, $\texttt{13a}_{4184}$ (each either 1 or 2), $\texttt{13n}_{65}$ (either 0 or 1), $\texttt{13n}_{459}$ (either 1 or 2).

1.6.1 Strong/Weak slice genus for links

For links, the notion of slice genus generalises in two different versions. Let L be a link, $L = \sqcup_{i=1}^{n} L_i$.

Definition 1.19 (Weak slice genus) A *(weak) slice surface* for L is a compact, oriented surface $F \subset B^4$, such that $\partial F = L$.

The *(weak) slice genus* $g_4(L)$ is the minimal genus of a (weak) slice surface for L.

Definition 1.20 (Strong slice genus) A *strong slice surface* for L is a compact, oriented surface $F = \sqcup_{i=1}^{n} F_i \subset B^4$, such that $\partial F_i = L_i$.

The *strong slice genus* $g_4^*(L)$ is the minimal genus of a strong slice surface for L.

Remark that we can also require weak slice surfaces to be *connected* as well: if we get two disjoint surfaces, we can perform a final band addition between the two unknots we obtained. This results in a connected surface and it does not increase the genus of the surface we built so far.

On the other hand, strong slice surfaces are necessarily the disjoint union of several surfaces: by definition, they are the collection of "one surface per link component".

Remark moreover that only links with linking numbers zero can admit a strong slice surface (see [Cav15, Section 5.1]).

1.7 Bounds on the unknotting number and on the slice genus

While upper bounds often come from explicit constructions (e.g. Seifert surfaces, braid techniques, direct unknotting/slicing sequences), there are some lower bounds on the values of the unknotting number and the slice genus coming from theoretical results.

Remark first that $g_4(K) \leq u(K)$: each crossing change can be realised by a single saddle move on a cobordism, so a sequence of $u(K)$ crossing changes from K to the unknot gives a genus-$u(K)$ surface in B^4. Hence, any lower bound for the slice genus is also a lower bound for the unknotting number.

Lower bounds are typically obtained via inequalities with other topological invariants. A non-exhaustive list is the following:

- From the τ-invariant, obtained via Heegaard Floer Homology, we get the inequality $|\tau(K)| \leq g_4(K) \leq u(K)$ [OS03];
- From the Rasmussen s-invariant, obtained via Heegaard Floer Homology, we get the inequality $\frac{1}{2}|s(K)| \leq g_4(K) \leq u(K)$ [Ras10];

- From the signature, computed directly on a diagram of K, we get the inequality $\frac{1}{2}\,|\sigma(K)| \leq g_4(K) \leq u(K)$.

Moreover, exact formulas are known for some families, e.g. torus knots, for which the Milnor conjecture gives

$$g_4(T(p,q)) = \frac{(p-1)(q-1)}{2} = u(T(p,q));$$

for a proof, one can see [Ras10, Corollary 1].

Despite all these tools, exact values often remain out of reach: even for the small knots listed above, the lower and upper bounds simply do not match, and no known invariant pins down $g_4(K)$. This is again one of the motivations for our use of RL: a sequence of band paths found by the agent that realises a low-genus slice surface provides an upper bound that may close the gap with the known lower bounds.

2 Reinforcement Learning for Knot Theory

The application of machine learning to knot-theoretic problems has developed rapidly over the last decade. On the supervised side, neural networks have been trained to predict knot invariants directly from a diagram [Hug20] and to recover the hyperbolic volume of a knot from its Jones polynomial [JKP19]; more broadly, machine-learning attribution techniques have been used to surface previously unknown relationships between invariants, leading to new conjectures in knot theory [Dav+21]. On the reinforcement learning side, several works have framed the simplification of a knot as a sequential decision problem: Gukov, Halverson, Ruehle and Sułkowski [Guk+21] trained an agent to recognise and untie the unknot via Markov moves and braid relations, while Applebaum, Blackwell, Davies, Edlich, Juhász, Lackenby, Tomašev and Zheng [App+25] used a reinforcement-learning agent to bound the unknotting number, producing along the way a large dataset of hard unknot diagrams.

Closer to our work, Gukov, Halverson, Manolescu and Ruehle [Guk+25] searched for ribbon bands exhibiting knots as slice: they cast the search for a ribbon disc (a particular genus-zero surface) as a band-path game very close to the one we described in Section 2.2. The present work follows this line, unifying the unknotting number and the more general slice genus into a single band-path game.

2.1 RL basics

Reinforcement Learning (RL) is a Machine Learning framework for sequential decision problems. As Sutton and Barto state in [SB98],

> Of all the forms of machine learning, reinforcement learning is the closest to the kind of learning that humans and other animals do, and many of the core algorithms of reinforcement learning were originally inspired by biological learning systems.

The core idea of RL, learning from rewards and punishments, goes back to the 1950s, but its most striking demonstrations are recent: *TD-Gammon* reached near world-class backgammon in the 1990s, and between 2016 and 2018 *AlphaGo* and *AlphaGo Zero* mastered Go, the latter learning entirely from self-play, without ever looking at a human game.

The basic picture is an *agent* that interacts with an environment in a loop (Figure 17): from the current state it considers several candidate actions, each is applied, the resulting states are evaluated and for each a reward is returned; the best action is kept as the next move, and the process repeats until the episode ends. Through trial and error, an initially random agent gradually learns to favour the actions that lead to higher cumulative reward.

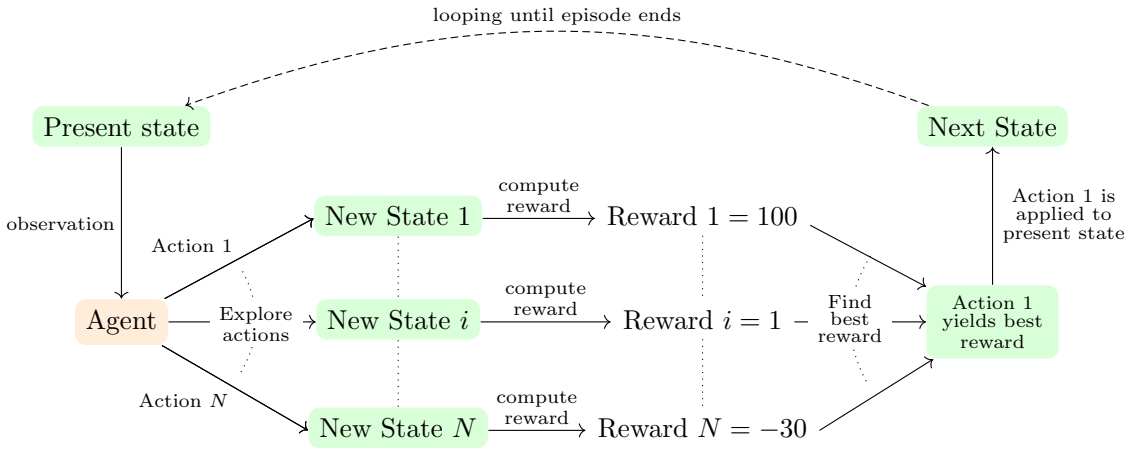


Figure 17: Schematic of the RL loop.

More precisely, the agent's strategy is a *policy* $\pi(a \mid s)$, a typically probabilistic map assigning to each state s a distribution over actions. Its aim is not to maximise the immediate reward but the *expected cumulative reward*

$$\mathbb{E}_\pi \left[\sum_{t \geq 0} \gamma^t \, r_t \right]$$

collected over an episode, where r_t is the reward at time t and $\gamma \in (0, 1]$ discounts future rewards. To this end it also estimates a *value function* $V^\pi(s)$, the expected cumulative reward obtained starting from s and following π thereafter.

A core challenge in reinforcement learning is the *exploration–exploitation tradeoff*: the agent must *exploit* the actions that currently look best while still *exploring* less familiar ones, to avoid getting stuck in suboptimal habits. The policy is not given in advance: it is *learned* over many episodes. In the *deep reinforcement learning* setting used in this work, π and V^π are represented by *neural networks* and updated via gradient-based optimisation. This is the approach that powered the AlphaGo programs and underlies the work described in the rest of this report.

Example 2.1 (Tetris) Tetris fits the RL framework naturally: an *episode* is one game, a *state* is the board together with the falling piece, an *action* is one of move/rotate/drop, and the *reward* is positive when a row is cleared and strongly negative when the stack reaches the top.

2.2 Using RL to compute unknotting number and slice genus

We can phrase the computation of the unknotting number and the slice genus as a game. The RL framework is set up as follows:

- **Episode**: starting from a link and trying to reach the unlink.
- **State**: link diagram and the band paths already chosen.
- **Action**: choose a new band path.
- **Reward**: the smaller the resulting unknotting number (or slice genus), the higher the score.

On the implementation side, the link diagram is encoded as the data of

- the vector of the PD code;
- the vector of crossing signs;
- a handful of bookkeeping scalars (number of link components, of unknotted unlinked components, of connected components).

Together these uniquely determine the oriented diagram. During an episode, the observation state is enriched with the previous band paths and the running one.
To this raw combinatorial encoding, we concatenate a set of topological invariants, computed after every band path; this includes the Seifert genus, the Heegaard Floer invariants ν and τ, the linking matrix, the determinant and signature.
The current state of the episode is then fed to a pair of neural networks: a policy network that outputs a distribution over candidate band paths, and a value network that estimates the expected cumulative reward of the current state. Both networks are built on top of Stable-Baselines3's MultiInputPolicy. The training algorithm we use is Proximal Policy Optimization (PPO), an on-policy actor–critic policy-gradient method introduced by Schulman, Wolski, Dhariwal, Radford and Klimov in 2017 [Sch+17], currently among the most widely used RL algorithms in practice. We train for roughly 200,000 timesteps to obtain a well-performing agent.

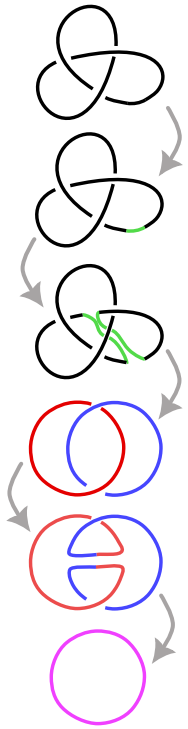


Figure 18: History of band additions in the RL framework.

2.3 Results

Using the band-path game and the reinforcement-learning agent described above, we obtained new upper bounds on the unknotting number and on the (weak and strong) slice genus for several families of knots and links. Most of these bounds were produced directly by the trained agent; a few were obtained, or sharpened, by combining the agent's upper bounds with the classical lower bounds of Section 1.7. We summarise them family by family.

2.3.1 Unknotting number of links with 11 crossings

On [KnotInfo] it is possible to find the unknotting number of prime links with up to 10 crossings: these were computed by Bulai in [Bul18]. Out of the 417 unique prime links with up to 10 crossings, the unknotting number is known for all but two: these are $\texttt{L10n}_{32}$ and $\texttt{L10n}_{34}$.

To the best of our knowledge, no table of unknotting numbers for multi-component prime links with 11 crossings was previously available. Disregarding orientations, there are 1,007 unique prime links with 11 crossings: the agent determined the exact unknotting number for 321 of them, and confined each of the remaining 686 to two consecutive integers.

2.3.2 Slice genus of knots with 14 crossings

There are 12,965 prime knots with up to 13 crossings. For all but 19, it is possible to find the exact slice genus online (for instance, see [KnotInfo]).

No slice genus dataset for 14-crossing prime knots was available online. We ran a full census of the 46,972 knots from scratch. We obtained the *exact* slice genus for 32,288 of them and, for every one of the remaining 14,684, narrowed it down to two consecutive integers.

2.3.3 Weak slice genus of links with up to 11 crossings

Among the 4,188 prime links with up to 11 crossings, the weak slice genus was previously unknown for 407 of them. We determined it for 125 of these, all of which turn out to be slice (weak slice genus 0). The updated values have been submitted to [LinkInfo] and can be found there.

2.3.4 Strong slice genus of links with up to 11 crossings

As remarked in Section 1.6.1, the strong slice genus is only defined for links whose pairwise linking numbers vanish. There are 732 prime links with up to 11 crossings that have this property and for which we can compute the strong slice genus. We were not able to find any database online, and we carried out a full census of them. The agent pinned down the *exact* value for 374 of them, and constrained each of the remaining 358 to an interval of consecutive integers.

References

[KnotInfo] Charles Livingston and Allison H. Moore, "KnotInfo: Table of Knot Invariants". url: `https://knotinfo.org/`.

[LinkInfo] Charles Livingston and Allison H. Moore, "LinkInfo: Table of Link Invariants". url: `https://knotinfo.org/linkinfo/`.

[Ada04] Colin C. Adams, "The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots". Providence, RI: American Mathematical Society, 2004. isbn: 978- 0-8218-3678-1.

[App+25] Taylor Applebaum et al., *The unknotting number, hard unknot diagrams, and reinforcement learning.* In: Experimental Mathematics (2025). doi: 10.1080/ 10586458.2025.2542174. arXiv: 2409.09032 [math.GT].

[Bul18] Lavinia Bulai, *Unlinking numbers of links with crossing number 10.* In: Involve 11.2 (2018), pp. 335–353. doi: 10.2140/involve.2018.11.335.

[Cav15] Alberto Cavallo, *On the slice genus and some concordance invariants of links.* In: Journal of Knot Theory and Its Ramifications 24.4 (2015), p. 1550021. doi: 10.1142/S0218216515500212.

[CL14] Alexander Coward and Marc Lackenby, *An upper bound on Reidemeister moves.* In: American Journal of Mathematics 136.4 (2014), pp. 1023–1066. doi: 10.1353/ ajm.2014.0027. arXiv: 1104.1882 [math.GT].

[Cro04] Peter R. Cromwell, "Knots and Links". Cambridge: Cambridge University Press, 2004.

[Dav+21] Alex Davies et al., *Advancing mathematics by guiding human intuition with AI.* In: Nature 600.7887 (2021), pp. 70–74. doi: 10.1038/s41586-021-04086-x.

[Guk+21] Sergei Gukov et al., *Learning to unknot.* In: Machine Learning: Science and Tech- nology 2.2 (2021), p. 025035. doi: 10.1088/2632-2153/abe91f. arXiv: 2010.16263 [math.GT].

[Guk+25] Sergei Gukov et al., *Searching for ribbons with machine learning.* In: Machine Learning: Science and Technology 6.2 (2025), p. 025065. doi: 10.1088/2632- 2153/ade362. arXiv: 2304.09304 [math.GT].

[Hug20] Mark C. Hughes, *A neural network approach to predicting and computing knot invariants.* In: Journal of Knot Theory and Its Ramifications 29.3 (2020), p. 2050005. doi: 10.1142/S02182165 20500054. arXiv: 1610.05744 [math.GT].

[JKP19] Vishnu Jejjala, Arjun Kar, and Onkar Parrikar, *Deep learning the hyperbolic vol- ume of a knot.* In: Physics Letters B 799 (2019), p. 135033. doi: 10.1016/j. physletb.2019.135033. arXiv: 1902.05547 [hep-th].

[Lac15] Marc Lackenby, *A polynomial upper bound on Reidemeister moves.* In: Annals of Mathematics 182.2 (2015), pp. 491–564. doi: 10.4007/annals.2015.182.2.3. arXiv: 1302.0180 [math.GT].

[OS03] Peter Ozsváth and Zoltán Szabó, *Knot Floer homology and the four-ball genus.* In: Geometry & Topology 7.2 (2003), pp. 615–639. doi: 10.2140/gt.2003.7.615.

[Ras10] Jacob Rasmussen, *Khovanov homology and the slice genus.* In: Inventiones Mathematicae 182.2 (2010), pp. 419–447. doi: 10.1007/s00222-010-0275-6.

[SB98] R.S. Sutton and A.G. Barto, "Reinforcement Learning: An Introduction". A Bradford book. MIT Press, 1998. isbn: 9780262193986. url: `https://books.google.it/books?id=CAFR6IBF4xYC`.

[Sch+17] John Schulman et al., "Proximal Policy Optimization Algorithms". arXiv: 1707. 06347 [cs.LG].

[Sei35] Herbert Seifert, *Über das Geschlecht von Knoten.* In: Mathematische Annalen 110 (1935), pp. 571–592. doi: 10.1007/BF01448044.

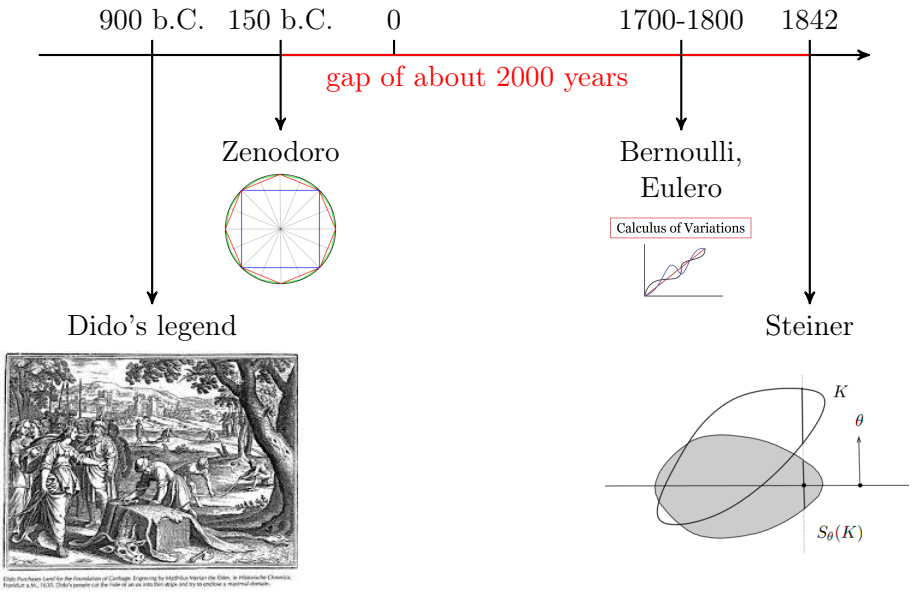
Isoperimetric problem in the Heisenberg group

Davide Massenz [(*)]

Abstract. The isoperimetric problem asks which sets minimize perimeter among all sets with a fixed volume. This is one of the most ancient questions in mathematics, dating back to the legend of Queen Dido. Despite the ancient origins, a turning point came with the work of De Giorgi only around the 1950s. He introduced a general definition of perimeter in the n-dimensional Euclidean space and this allowed for a more rigorous formulation of the isoperimetric problem. Later the problem has been generalized by other mathematicians to different frameworks, like Riemannian manifolds and metric spaces.
In this talk, after an historical introduction, we will at first recall the well-known isoperimetric inequality in the euclidean case, where the solution is the ball, examining in detail the notion of perimeter on which it is based. We will then introduce the Heisenberg group as the simplest and most studied example of a sub-Riemannian manifold and we will discuss the isoperimetric problem in this setting, highlighting Pansu's conjecture which provides the candidate isoperimetric sets, known as Pansu spheres.

[(*)]Ph.D. course, Università di Padova, Dip. Matematica, via Trieste 63, I-35121 Padova, Italy. E-mail: massenz@math.unipd.it . Seminar held on 20 November 2025.

1 Historical Introduction



The earliest scientific work on the isoperimetric problem for which reliable historical evidence exists is the lost treatise "On Isoperimetric Figures" by Zenodorus. This work reportedly contained, among other results, proofs of the isoperimetric properties of regular polygons and of the circle [1]. Prior to Zenodorus, the only accounts related to the problem belong to the legendary tradition associated with Queen Dido. However, this tradition is not supported by evidence of any philosophical or scientific treatises predating Zenodorus' work.

According to the myth transmitted most famously in Virgil's Aeneid, Dido, queen of Tyre, fled to North Africa after being exiled by her brother Pygmalion. There she negotiated with a local ruler to obtain as much land as could be enclosed by a single oxhide. By cutting the hide into thin strips and tying them together to form a long rope, she enclosed a region along the coast, which, according to the legend, became the site of Carthage. The mathematical question implicit in this story, now known as Dido's problem, is to determine the shape of a closed curve of given length that encloses the maximum possible area. The answer suggested by the legend, namely the circumference, is indeed the correct one.

After antiquity, sustained mathematical interest in isoperimetric questions re-emerged only with the development of the calculus of variations. At the end of the seventeenth century, the Bernoulli brothers formulated the brachistochrone problem, and in 1738 Euler laid the foundations of the modern calculus of variations. Nevertheless, the first modern geometric treatment of the classical planar isoperimetric problem appeared only in 1842, when Jakob Steiner published his solution based on symmetrization arguments (assuming existence).

Thus, the first substantial generalization and rigorous development of the geometric ideas already present in Zenodorus' treatise was achieved nearly two millennia later.

2 Notion of Perimeter

The long delay in the rigorous treatment of the isoperimetric problem is largely due to the intrinsic difficulty of defining surface area in a sufficiently general framework, namely, for arbitrary Lebesgue measurable subsets of $\mathbb{R}^n$.

2.1 Perimeter in $\mathbb{R}^2$

In the plane, the notion of perimeter is classical and relatively straightforward.

Definition 2.1 (Perimeter in $\mathbb{R}^2$) Let $\gamma : [a, b] \to \mathbb{R}^n$ be a continuous curve. For any finite partition of $[a, b]$ $\mathcal{P} = \{a = t_0 < t_1 < \cdots < t_m = b\}$, define the polygonal approximation length

$$L(\gamma, \mathcal{P}) = \sum_{i=1}^{m} \|\gamma(t_i) - \gamma(t_{i-1})\|.$$

The length of the curve γ is then defined as

$$L(\gamma) = \sup_{\mathcal{P}} L(\gamma, \mathcal{P}) = \sup_{\mathcal{P}} \sum_{i=1}^{m} \|\gamma(t_i) - \gamma(t_{i-1})\|,$$

where the supremum is taken over all finite partitions $\mathcal{P}$ of $[a, b]$.

For piecewise C^1 curves this definition coincides with the classical formula

$$L(\gamma) = \int_a^b |\gamma'(t)| dt.$$

In higher dimensions, however, the situation becomes considerably more delicate. Approximating a surface by polyhedral domains (a set that is contained in the union of a finite number of hyperplanes) may lead to divergent areas even when the original surface has finite area. This phenomenon was famously demonstrated by Schwarz in 1890 [2].

2.2 Hausdorff measure

A first candidate for surface area known since 1920s is the Hausdorff measure.

Definition 2.2 (Hausdorff measure) Let $A \subseteq \mathbb{R}^n$ and $k \in \mathbb{N}$. For $\delta > 0$, define the δ-approximated k-dimensional Hausdorff outer measure of A by

$$\mathcal{H}^k_\delta(A) = \frac{\omega_k}{2^k} \inf \left\{ \sum_{i=1}^{\infty} (\operatorname{diam} U_i)^k : A \subseteq \bigcup_{i=1}^{\infty} U_i,\ \operatorname{diam}(U_i) < \delta \right\},$$

where $\omega_k = \mathcal{L}^k(B^k(0, 1)) = \frac{\pi^{k/2}}{\Gamma(1+k/2)}$ and $\operatorname{diam}(A) = \sup_{x,y \in A} |x - y|$. The k-dimensional Hausdorff measure of A is then defined as

$$\mathcal{H}^k(A) = \lim_{\delta \to 0} \mathcal{H}^k_\delta(A) = \sup_{\delta > 0} \mathcal{H}^k_\delta(A).$$

The Hausdorff measure is not well suited for variational problems, since it fails to be lower semicontinuous with respect to Hausdorff convergence of sets. Recall that the Hausdorff distance is defined by

$$d_H(A,B) = \max\left\{\sup_{x\in A}\operatorname{dist}(x,B), \sup_{y\in B}\operatorname{dist}(A,y)\right\}.$$

Example: Consider the interval $[0,1]$ and for $h \in \mathbb{N}$ divide it into h^2 intervals and define A_h to be the union of the intervals which correspond to the integer multiples of h. The Hausdorff distance tends to 0

$$d_H(A_j,[0,1]) = \frac{n-1}{n^2} \to 0,$$

but

$$\liminf_h \mathcal{H}^1(A_h) = \lim_h \frac{1}{h} = 0 < 1 = \mathcal{H}^1(A).$$

Thus lower semicontinuity fails.

2.3 Cacciopoli's perimeter

In 1953, Renato Caccioppoli introduces the following definition of surface area of a 2-dimensional manifold in $\mathbb{R}^3$ [3].

Definition 2.3 (Caccioppoli's perimeter) The perimeter of a Lebesgue measurable set $E \subset \mathbb{R}^3$ is

$$\inf\left\{\liminf_{j\to\infty} P(\Pi_j) : \mathcal{L}^3(\Pi_j \Delta E) \underset{j\to\infty}{\longrightarrow} 0\right\},$$

where Π_j are polyhedral domains and $\Pi_j \Delta E$ denotes the symmetric difference $\Pi_j \Delta E = (\Pi_j \backslash E) \cup (E \backslash \Pi_j)$.

This definition certainly has a straightforward geometric interpretation, but it is not very practical. In 1954, E. De Giorgi proposed a more analytical definition of perimeter, which he proved to be equivalent and which has become the standard definition of perimeter [4].

2.4 De Giorgi's perimeter

Definition 2.4 (De Giorgi's perimeter) Let $E \subset \mathbb{R}^n$ be a Lebesgue measurable set. The perimeter of E is

$$P(E) = \sup\left\{\int_E \operatorname{div}\varphi(x)dx : \varphi \in C_c^1(\mathbb{R}^n,\mathbb{R}^n), \sup_{\mathbb{R}^n}|\varphi| \leq 1\right\}.$$

The idea behind this definition is that it is a generalization of the Gauss-Green formula: given $\varphi \in C^1(\mathbb{R}^n)$ and a set $E \subset \mathbb{R}^n$ with C^1-boundary, by the Gauss-Green theorem

$$\int_E \operatorname{div} \varphi dx = \int_{\partial E} \varphi \cdot \nu_E d\mathcal{H}^{n-1},$$

where ν^E denotes the outer unit normal to the boundary of E. Assuming $|\varphi| \leq 1$, we hence get

$$\int_E \operatorname{div} \varphi dx \leq \mathcal{H}^{n-1}(\partial E),$$

and choosing $\varphi = \nu_E$ we get $P(E) = \mathcal{H}^{n-1}(\partial E)$ for sets with smooth boundary.

With this defintion the lower semi-continuity is straightforward.

Theorem 2.5 (Lower semi-continuity) *Let $E, E_k \subset \mathbb{R}^n$ measurable sets such that $\chi_{E_k} \to \chi_E$ in $L^1_{\text{loc}}(\mathbb{R}^n)$, then*

$$P(E) \leq \liminf_{k\to\infty} P(E_k).$$

Proof. For each fixed $\varphi \in C^1_c(\mathbb{R}^n, \mathbb{R}^n)$ with $\sup|\varphi| \leq 1$, by dominated convergence since $\chi_{E_k} \to \chi_E$ in $L^1_{\text{loc}}(\mathbb{R}^n)$ we have that

$$\int_E \operatorname{div} \varphi \, dx = \lim_{k\to\infty} \int_{E_k} \operatorname{div} \varphi \, dx \leq \liminf_{k\to\infty} P(E_k).$$

Taking the supremum over all φ gives $P(E, \Omega) \leq \liminf_{k\to\infty} P(E_k, \Omega)$. □

De Giorgi's formulation thus provides the natural framework for geometric variational problems and forms the foundation of the modern theory of sets of finite perimeter.

3 Isoperimetric inequality in $\mathbb{R}^n$

The notion of perimeter introduced by Caccioppoli and De Giorgi provides the appropriate framework for the modern formulation of the isoperimetric problem in $\mathbb{R}^n$.

Theorem 3.1 (Isoperimetric inequality in $\mathbb{R}^n$) *For any Lebesgue measurable set $E \subset \mathbb{R}^n$ with finite measure we have*

$$P(E) \geq n\omega_n^{\frac{1}{n}} \mathcal{L}^n(E)^{\frac{n-1}{n}}$$

Equality case occurs if and only if E is a Euclidean ball.

This theorem states that the ball has the least perimeter among all sets with the same measure, and it is the unique set having this property.

The proof of the existence of isoperimetric sets is based on the two following properties of the perimeter:

- Lower semi-continuity,

- compactness: let $(E_h)_{h\in\mathbb{N}} \subset \mathbb{R}^n$ be a sequence of sets with finite perimeter such that

$$\sup_{h\in\mathbb{N}} P(E_h) < \infty,$$

then there exists a subsequence and a set of finite perimeter $E \subset \mathbb{R}^n$ such that $\mathcal{L}^n(E \triangle E_{h_k}) \to 0$.

Once we have these two properties, the conclusion is trivial. Consider a minimizing sequence $(E_h)_{h\in\mathbb{N}}$ of $\inf\{P(E) : \mathcal{L}^n(E) = v\} =: m$, i.e. $\mathcal{L}^n(E_h) = v \ \forall h \in \mathbb{N}$ and $\lim P(E_h) = m$. Then by compactness there exists a subsequence and a set of finite perimeter $E \subset \mathbb{R}^n$ such that $\mathcal{L}^n(E \triangle E_{h_k}) \to 0$ and by lower semi-continuity $P(E) \leq \liminf P(E_{h_k}) = m$ and hence the minimum is attained.

The uniqueness and the characterization of the minimizer is obtained through classical symmetrization techniques,like Steiner symmetrization. The idea of Steiner rearrangement is the following. Let $E \subset \mathbb{R}^n$ be a measurable set and fix a direction, say the x_n-axis. The Steiner symmetrization of E with respect to the hyperplane $\{x_n = 0\}$ is obtained by replacing, for every $x' \in \mathbb{R}^{n-1}$, the one-dimensional slice

$$E_{x'} = \{\, t \in \mathbb{R} : (x', t) \in E \,\}$$

with the centered interval in the x_n-direction having the same one-dimensional Lebesgue measure as $E_{x'}$. The resulting set is symmetric with respect to $\{x_n = 0\}$ and has the same volume of E. Moreover, it can be proved that the symmetrization does not increase the perimeter. Iterating this procedure in all directions leads to a ball, showing that balls uniquely minimize perimeter among sets of fixed volume.

4 Heisenberg group

The Heisenberg group $\mathbb{H}^1$ is the Lie group $(\mathbb{R}^3, *)$, where the product $*$ is defined, for any pair of points $[z,t], [z',t'] \in \mathbb{R}^3 \equiv \mathbb{C} \times \mathbb{R}$, as

$$[z,t] * [z',t'] := [z + z', t + t' + \mathrm{Im}(z\bar{z}')] \quad (z = x + iy).$$

This non-commutative structure endows $\mathbb{H}^1$ with the geometry of a step-two nilpotent Lie group. The Lie algebra of $\mathbb{H}^1$ is generated by the following basis of left invariant vector fields:

$$X := \frac{\partial}{\partial x} + y\frac{\partial}{\partial t}, \quad Y := \frac{\partial}{\partial y} - x\frac{\partial}{\partial t}, \quad T := \frac{\partial}{\partial t}.$$

Definition 4.1 (Horizontal distribution) The horizontal distribution $\mathcal{H}$ in $\mathbb{H}^1$ is the smooth planar one generated by X and Y:

$$\mathcal{H}_p = \mathrm{span}\{X_p, Y_p\} \quad \text{for } p \in \mathbb{H}^1.$$

The horizontal projection of a vector U onto $\mathcal{H}$ will be denoted by U_H. A vector field U is called horizontal if $U = U_H$. A horizontal curve is a C^1 curve whose tangent vector lies in the horizontal distribution.

They satisfy the commutation relations

$$[X,T] = [Y,T] = 0, \qquad [X,Y] = -2T,$$

which show that the distribution generated by X and Y is bracket generating. We shall consider on $\mathbb{H}^1$ the (left invariant) Riemannian metric $g = \langle \cdot, \cdot \rangle$ so that $\{X, Y, T\}$ is an orthonormal basis at every point, and the associated Levi-Civita connection D. The modulus of a vector field U will be denoted by $|U|$.

Definition 4.2 (Carnot-Carathédory distance) The Carnot-Carathéodory distance d_{cc} between two points in $\mathbb{H}^1$ is defined as the infimum of the length of horizontal curves joining the given points:

$$d(x,y) = \inf \left\{ \int_0^1 \sqrt{\langle \gamma', X|_{\gamma(t)} \rangle^2 + \langle \gamma', Y|_{\gamma(t)} \rangle^2} dt : \gamma'(t) \in \mathcal{H}_{\gamma(t)}, \gamma(0) = x, \gamma(1) = y \right\}.$$

The Heisenberg group admits a one-parameter family of anisotropic dilations generated by the vector field

$$W = xX + yY + 2tT.$$

In coordinates,

$$\varphi_s(x,y,t) = (e^s x,\, e^s y,\, e^{2s} t).$$

These dilations reflect the stratified structure of the Lie algebra: horizontal directions scale linearly, while the vertical direction scales quadratically.

4.1 Volume and Perimeter in $\mathbb{H}^1$

Now we introduce the notions of volume and area in $\mathbb{H}^1$.

- The volume $V(E)$ of a Borel set $E \subseteq \mathbb{H}^1$ is the Riemannian volume of the left invariant metric g, which coincides with the Lebesgue measure in $\mathbb{R}^3$.
- Given a $E \in \mathbb{H}^1$ measurable set, the (horizontal) perimeter is defined as

  $$P_{\mathbb{H}}(E) = \sup \left\{ \int_E \mathrm{div}_{\mathbb{H}}\, \varphi(x) dx : \varphi \in C_c^1(\mathbb{R}^3, \mathbb{R}^2), \sup_{\mathbb{R}^3} |\varphi| \leq 1 \right\},$$

  where $\mathrm{div}_{\mathbb{H}}\, \varphi(x) = X\varphi_1(x) + Y\varphi_2(x)$. Again, this perimeter is lower semi-continuous with respect to L^1_{loc} convergence. If E has a smooth boundary Σ with a unit normal vector field N it is equivalent to

  $$A(\Sigma) := \int_\Sigma |N_H|\, d\Sigma.$$

  where $N_H = N - \langle N, T \rangle T$, and $d\Sigma$ is the Riemannian area element on Σ.

Any isometry of $(\mathbb{H}^1, g)$ leaving invariant the horizontal distribution preserves the area of surfaces in $\mathbb{H}^1$. Examples of such isometries are left translations, which act transitively on $\mathbb{H}^1$ and the Euclidean rotation of angle θ about the t-axis. The dilations changes the volume and the perimmeter in the following way

$$V(\lambda E) = \lambda^4 V(E), \quad P_{\mathbb{H}}(\lambda E) = \lambda^3 P_{\mathbb{H}}(E).$$

4.2 Geodesics and Pansu sphere

Geodesics in $\mathbb{H}^1$ are defined as the horizontal length minimizing curves joining pairs of points. We may think of a geodesic in $\mathbb{H}^1$ as a smooth horizontal curve that is a critical point of length under any variation by horizontal curves with fixed endpoints. Indeed we have the following theorem.

Theorem 4.3 (Geodesic equation) *Let $\gamma : I \to \mathbb{H}^1$ be a C^2 horizontal curve parameterized by arc-length. Then γ is a critical point of length for any admissible variation if and only if there is $\lambda \in \mathbb{R}$ such that γ satisfies the second order ordinary differential equation*

$$D_{\dot\gamma}\dot\gamma + 2\lambda J(\dot\gamma) = 0, \tag{G}$$

where $J(U) := D_U T$ for any vector field U. It's easy to see that $J(X) = Y, J(Y) = -X$ and $J(T) = 0$, so that $J^2 = -Id$ when restricted to the horizontal distribution.

Then we define the geodesics in the following way.

Definition 4.4 (Geodesics) We will say that a C^2 horizontal curve γ is a geodesic of curvature λ if it is parameterized by arc-length and satisfies eq. (G).

Let us see how these geodesics look like. Due to left-invariance of the metric, we may assume that the two points are $(0,0,0)$ and $p = (x, y, t)$. Let $\gamma : [0,1] \to \mathbb{H}^1$ be a horizontal curve joining 0 and p. Let $\tilde\gamma$ be the closed curve obtained by closing $\pi(\gamma)$ with the segment joining 0 and $\pi(p)$ and S be the region enclosed by $\tilde\gamma$. By Stoke's theorem we have

$$t = \int_0^1 \gamma_3' dt = \frac{1}{2}\int_0^1 (\gamma_1\gamma_2' - \gamma_2\gamma_1')dt = \frac{1}{2}\int_{\tilde\gamma} x dy - y dx = \int_S dx \wedge dy = \text{Area}(S).$$

Then the problem of finding the horizontal length minimizing curve between 0 and p is equivalent to find the plane curve from 0 to (x, y) with minimum length, subject to the constraint that the region S delimited by the curve and the segment joining 0 and (x, y) has fixed area. This is Dido's problem and is solved by choosing the plane curve to be an arc of a circle. In conclusion, the geodesic between 0 and p is the lift of a circular arc joining 0 and (x, y), whose convex hull has area t.

From equation (G), we can compute explicitly the equations of a geodesic $\gamma(s) = (x(s), y(s), t(s))$ of curvature $\lambda \neq 0$ with initial conditions $(x_0, y_0, t_0) = (x(0), y(0), t(0))$ and $(A, B) = (\dot x(0), \dot y(0))$:

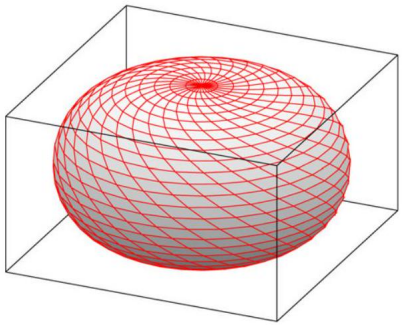
$$\begin{cases} x(s) = x_0 + A\left(\frac{\sin(2\lambda s)}{2\lambda}\right) + B\left(\frac{1-\cos(2\lambda s)}{2\lambda}\right) \\ y(s) = y_0 - A\left(\frac{1-\cos(2\lambda s)}{2\lambda}\right) + B\left(\frac{\sin(2\lambda s)}{2\lambda}\right) \\ t(s) = t_0 + \frac{1}{2\lambda}\left(s - \frac{\sin(2\lambda s)}{2\lambda}\right) + (Ax_0 + By_0)\left(\frac{1-\cos(2\lambda s)}{2\lambda}\right) - (Bx_0 - Ay_0)\left(\frac{\sin(2\lambda s)}{2\lambda}\right). \end{cases}$$

When $\lambda = 0$, we have $\gamma(t) = (x_0 + At, y_0 + Bt, t_0 + (Ay_0 - Bx_0)t)$.

Let us define now the Pansu sphere. Given $\lambda > 0$, we define $\mathbb{S}_\lambda$ as the union of all geodesics starting from 0 with curvature λ restricted to the interval $[0, \pi/\lambda]$. It is not difficult to see that $\mathbb{S}_\lambda$ is a compact embedded surface of revolution homeomorphic to a

sphere. Alternatively $\mathbb{S}_\lambda$ can be described as the union of the following radial graphs over the xy-plane

$$t = \frac{\pi}{4\lambda^2} \pm \frac{1}{2\lambda^2}\left(\lambda\rho\sqrt{1-\lambda^2\rho^2} + \arccos(\lambda\rho)\right), \quad \rho = \sqrt{x^2+y^2} \leqslant \frac{1}{\lambda}.$$



5 Isoperimetric inequality in $\mathbb{H}^1$

Now we can state the conjecture about the isoperimetric problem formulated by Pansu in 1982, see [5]. Let $\mathbb{B}$ be the Pansu ball, i.e. the set enclosed by $\mathbb{S}_1$.

Conjecture (Pansu's conjecture) *For any Lebesgue measurable set $E \subset \mathbb{H}^1$ with finite measure we have*

$$P_{\mathbb{H}}(E) \geq \frac{P_{\mathbb{H}}(\mathbb{B})}{V(\mathbb{B})^{3/4}} V(E)^{3/4} = \frac{4\sqrt{\pi}}{3^{3/4}} V(E)^{3/4},$$

and equality holds if and only if E is a Pansu ball (up to left translation and group dilation).

The Pansu's conjecture is still unsolved in this generality, although numerous partial results and special cases have been established over the years. For example, it has been proved assuming:

- C^2 smoothness [6],
- convexity [7],
- C^1 regularity and that it is given by an upper and lower graph over a disk [8].

For a detailed review on the Heisenberg isoperimetric problem we refer to the book [9].

6 Quantitative isoperimetric inequality

A different approach to the isoperimetric problem is given by Fuglede theorem, which proves in the euclidean case a quantitative isoperimetric inequality for nearly spherical graphs (see for example [10]).

A spherical graph is a set $E \subset \mathbb{R}^n$ whose boundary can be written as a graph over the boundary of the ball. So, if $u : \mathbb{S}^{n-1} \to \mathbb{R}$, the spherical graph associated to u is

$$E_u = \{y = tx(1+u(x)) \in \mathbb{R}^n : x \in \mathbb{S}^{n-1}, 0 \leq t < 1\}.$$

Theorem 6.1 (Fuglede theorem) *There exists $\varepsilon > 0$ and $c > 0$ such that for every nearly spherical graphs E_u with $\|u\|_{W^{1,\infty}} < \varepsilon$ and $V(E_u) = V(\mathbb{B}^n)$ and the barycenter of E_u is the origin, then*

$$P(E_u) - P(\mathbb{B}^n) \geq cV(E_u \triangle \mathbb{B}^n)^2.$$

We were able to prove the same inequality in the Riemannian approximation of the Heisenberg group (this result will appear soon in a work). The Riemannian approximation is the group $(\mathbb{H}^1, g_\varepsilon)$ where g_ε is the metric that makes orthonormal the following basis of left-invariant vector fields:

$$X := \frac{\partial}{\partial x} + y\frac{\partial}{\partial t}, \quad Y := \frac{\partial}{\partial y} - x\frac{\partial}{\partial t}, \quad T_\varepsilon := \varepsilon\frac{\partial}{\partial t}.$$

With the same construction seen before, one can define the candidate isoperimetric set $\mathbb{S}_\varepsilon$, with inner set $\mathbb{B}_\varepsilon$, which is an approximation of the Pansu sphere. Now consider a C^2 functions $u : \mathbb{S}_{\text{»}} \to \mathbb{R}$. Denote by S_{tu}, for t small, the immersed surface

$$S_{tu} = \left\{ \exp_p \left(tu(p) N_p \right) ; p \in \mathbb{S}_\varepsilon \right\},$$

where $\exp_p$ is the exponential map of $(\mathbb{H}^1, g_\varepsilon)$ at the point p and N_p is the Riemannian outer unit normal to $\mathbb{S}_\varepsilon$ at p. The family $\{S_{tu}\}$, for t small, is the variation of $\mathbb{S}_\varepsilon$ induced by u. Let E_{tu} be the region enclosed by S_{tu} and define

$$P(t) := P\left(E_{tu}\right), \quad V(t) := \mathcal{L}^3(E_{tu}).$$

Theorem 6.2 *There exists $\delta > 0$ small enough and $c > 0$ such that for any $u \in W^{1,2}(\mathbb{S}_\varepsilon)$ with $V(\Omega_u) = V(\mathbb{B}_\varepsilon)$, barycenter of Ω_u in the origin and $\|u\|_{C^{0,1}} < \delta$ it holds*

$$P(\Omega_u) - P(\mathbb{B}_\varepsilon) \geq cV(\Omega_u \triangle \mathbb{B}_\varepsilon)^2.$$

The argument follows the classical variational scheme. Consider the functional

$$\mathcal{F}(t) = P(t) - HV(t),$$

where H denotes the constant mean curvature of $\mathbb{S}_\varepsilon$. Expanding $\mathcal{F}$ at $t = 0$, the first variation vanishes because $\mathbb{S}_\varepsilon$ has constant mean curvature, while the second variation yields the Jacobi operator

$$\left.\frac{d^2}{dt^2}\mathcal{F}(t)\right|_{t=0} = J_\varepsilon(u) = \int_{\mathbb{S}_\varepsilon} \left(|\nabla_{\mathbb{S}_\varepsilon} u|^2 - (\mathrm{Ric}(N) + |h|^2)u^2 \right) d\mathbb{S}_\varepsilon.$$

The key point is the strict stability of $\mathbb{S}_\varepsilon$: under the volume and barycenter constraints, the quadratic form J_ε is strictly positive. This coercivity yields the desired quadratic control of the perimeter deficit.

The idea then is to pass to the limit $\varepsilon \to 0$ to obtain a formula like

$$P_{\mathbb{H}}(\Omega_u) - P_{\mathbb{H}}(\mathbb{B}) \geq cV(\Omega_u \triangle \mathbb{B})^2$$

under suitable assumption. This would provide a quantitative stability result for the isoperimetric problem in the Heisenberg group, paralleling Fuglede's theorem in the Euclidean case.

References

[1] Gian Paolo Leonardi et al., *Il mistero isoperimetrico di Zenodoro*. In Vedere la Matematica... alla maniera di Mimmo Luminati, p. 101–119. ETS, 2015.

[2] H.A. Schwarz, *Sur une définition erronée de l'aire d'une surface courbe*. In Gesammelte Mathematische Abhandlungen, p. 309–311. Berlin, 1890.

[3] Renato Caccioppoli, *Elementi di una teoria generale dell'integrazione k-dimensionale in uno spazio n-dimensionale*. In Atti del Quarto Congresso dell'Unione Matematica Italiana, Taormina, volume 2, pages 41–49, 1951.

[4] Ennio De Giorgi, *Su una teoria generale della misura* $(r-1)$*-dimensionale in uno spazio a* r *dimensioni*. Annali di Matematica Pura ed Applicata, 36(1): p. 191–213, 1954.

[5] Pierre Pansu, *An isoperimetric inequality on the Heisenberg-group*. Conference on differential geometry on homogeneous spaces (Turin, 1983). Rend. Sem. Mat. Univ. Politec. Torino, Special Issue, p. 159–174, 1983.

[6] Manuel Ritoré and César Rosales, *Area-stationary surfaces in the Heisenberg group* $\mathbb{H}^1$. Advances in Mathematics, 219(2): p. 633–671, 2008.

[7] Roberto Monti and Matthieu Rickly, *Convex isoperimetric sets in the Heisenberg group*. Annali della Scuola Normale Superiore di Pisa - Classe di Scienze, 8(2): p. 391–415, 2009.

[8] Manuel Ritoré, *A proof by calibration of an isoperimetric inequality in the Heisenberg group* $\mathbb{H}^n$. Calculus of Variations and Partial Differential Equations, 44(1-2): p. 47–60, 2012.

[9] Luca Capogna, Scott D. Pauls, and Donatella Danielli, "An introduction to the Heisenberg group and the sub-Riemannian isoperimetric problem". Springer, 2007.

[10] Nicola Fusco, *The quantitative isoperimetric inequality and related topics*. Bulletin of Mathematical Sciences, 5(3): p. 517–607, 2015.

Evolving geometries: an introduction to Mean Curvature Flow

Gaia Bombardieri (*)

Abstract. How do shapes evolve when driven by their mean curvature? In Euclidean space, Huisken's classical theorem ensures that convex surfaces become spherical as they shrink to a point. In this talk, we investigate the behavior of the flow in the sub-Riemannian setting of the Heisenberg group $\mathbb{H}^1$. We will first introduce the basics of classical Mean Curvature Flow, focusing on the role of self-shrinkers as models for singularities. Then, we will move to the sub-Riemannian context, analyzing the flow for mean convex hypersurfaces. We will discuss the problem of self-similarity in this setting and present a recent result: the Pansu sphere, despite being the candidate isoperimetric profile of $\mathbb{H}^1$, is not a self-shrinker. This reveals a striking divergence from the Euclidean intuition, where the static isoperimetric solution and the dynamic evolution profile coincide.

1 Curvature and mean convexity

1.1 The Second Fundamental Form and Mean Curvature

Let $\Sigma^n \hookrightarrow \mathbb{R}^{n+1}$ be a regular hypersurface. To describe its local geometry, we look at its *second fundamental form* A. Fixing an orthonormal basis $\{e_i\}_{i=1}^n$ of the tangent space $T_p\Sigma^n$, A can be represented at each point p as a symmetric $n \times n$ matrix with entries:

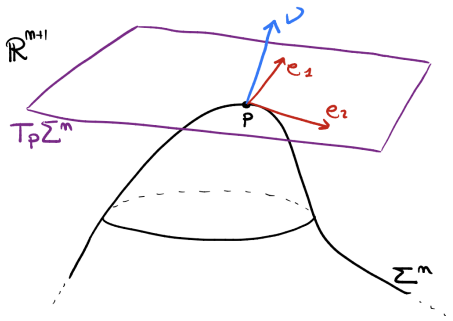
$$A_{ij}(p) = A(e_i, e_j)(p) = \langle \nabla_{e_i} \nu, e_j \rangle (p)$$

where $\nu : \Sigma \to \mathbb{S}^n$ is the external unit normal.

The *Mean Curvature* H is defined as the trace of the second fundamental form. It can be equivalently expressed as the sum of the principal curvatures $k_i(p)$, which are the eigenvalues of the second fundamental form (with respect to the induced metric) and describe the local geometry of the surface at the point p:

$$H(p) = \mathrm{tr}(A)(p) = \sum_{i=1}^{n} A_{ii}(p) = \sum_{i=1}^{n} k_i(p)$$

(*)Ph.D. course, Università di Padova, Dip. Matematica, via Trieste 63, I-35121 Padova, Italy. E-mail: bombardi@math.unipd.it . Seminar held on 17 December 2025.



1.2 Level set case

A particularly useful framework arises when the regular hypersurface is given as the level set of a smooth function $u : \mathbb{R}^{n+1} \to \mathbb{R}$. Specifically, if we define the surface as the boundary of a domain,

$$\Sigma = \partial\Omega = \{p \in \mathbb{R}^{n+1} \mid u(p) = c\}$$

for some constant $c \in \mathbb{R}$, then the geometric quantities can be expressed directly in terms of the derivatives of u. In this setting, the second fundamental form and the mean curvature are respectively given by:

$$A(e_i, e_j) = -\frac{\nabla\nabla u(e_i, e_j)}{|\nabla u|} \quad \text{and} \quad H = -\operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right),$$

where ∇u denotes the gradient of u and $\nabla\nabla u$ represents its Hessian matrix.

1.3 Convexity and Mean Convexity

Remark 1.1 Observe that $\partial\Omega$ has a positive semi-definite second fundamental form, i.e., $A \geq 0$, if and only if the domain Ω is convex.

Definition 1.2 A hypersurface Σ (or a domain Ω such that $\Sigma = \partial\Omega$) is said to be *mean convex* if its mean curvature is everywhere non-negative, that is, $H \geq 0$.

Figure 1: Examples of mean convex but not convex hypersurfaces.

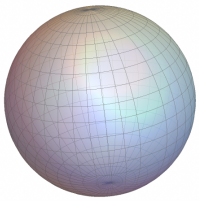
Remark 1.3 Since the mean curvature H is the trace of the second fundamental form A, it is algebraically immediate that every convex domain ($A \geq 0$) is also mean convex.

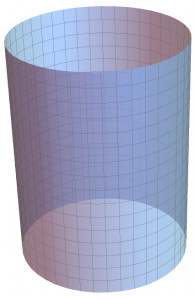
However, mean convexity is a strictly weaker geometric condition than standard convexity. For instance, thin tori or dumbbells can be constructed to be mean convex but are not convex.

Example 1.4 As a trivial example, for a hyperplane, the second fundamental form vanishes entirely, meaning $A = 0$ since the normal ν is constant.
Another classic example is the sphere of radius R, denoted as $\mathbb{S}^n_R$. If we consider $g_{\mathbb{S}^n_R}$ to be the round metric, i.e. the one induced by the Euclidean one, we have that

$$A(e_i, e_j) = \frac{1}{R} g_{\mathbb{S}^n_R}(e_i, e_j) = \frac{1}{R}\delta_{ij} \quad \Longrightarrow \quad H = \frac{n}{R}.$$

This formulation shows that the sphere is strictly convex.
Similarly, consider a vertical cylinder in $\mathbb{R}^{n+1}$, naturally defined as $\mathbb{S}^{n-1}_R \times \mathbb{R}$. The principal curvature corresponding to the flat vertical direction is 0, while the remaining $n-1$ principal curvatures along the spherical cross-section are $1/R$. Consequently, its mean curvature is strictly positive ($H = \frac{n-1}{R}$), demonstrating that the cylinder is mean convex ($H \geq 0$), but it is not strictly convex since one principal curvature is zero.





2 Mean Curvature Flow in Euclidean space

We say that a family of immersions $F_t : \Sigma^n \hookrightarrow \mathbb{R}^{n+1}$ evolves by Mean Curvature Flow (MCF) if it satisfies the following parabolic evolution equation:

$$\frac{\partial}{\partial t} F_t = -H_t \nu_t$$

where H_t and ν_t are respectively the mean curvature and the external unit normal of the hypersurface $\Sigma_t = F_t(\Sigma)$ at time t.

A fundamental reason to study MCF is that it represents the formal L^2-gradient flow of the area functional. Indeed, the area of the evolving surface is strictly decreasing over time, following the relation:

$$\frac{d}{dt}\text{Area}(\Sigma_t) = -\int_{\Sigma_t} H^2 \, d\mu \leq 0$$

A special and fundamentally important class of solutions to the Mean Curvature Flow is given by the so-called *self-shrinkers*. Geometrically, a self-shrinker is a hypersurface that evolves purely by homothetic dilations, preserving its overall shape while contracting, until it eventually collapses to a point in finite time.

2.1 Key Properties and Examples

The MCF exhibits several crucial geometric properties:

- **Curves in $\mathbb{R}^2$:** According to the classical curve shortening flow theorems, every smooth, simple closed curve in the plane eventually becomes convex and shrinks to a round point (see [9] and [11]).

- **The Round Sphere:** A round sphere $\Sigma_0 = \mathbb{S}^n_{R_0}$ (with $H = \frac{n}{R_0}$) homothetically shrinks to a point. The radius evolves exactly as $R(t) = \sqrt{R_0^2 - 2nt}$. The surface collapses to a point in finite time $T = \frac{R_0^2}{2n}$, acting as a perfect self-shrinker.
- **Maximum Principle:** The flow enjoys a strict maximum principle, also known as the avoidance principle [14, Chapter 2]. If two compact sets are initially contained within each other (or are disjoint), they will remain so under the evolution by MCF forever. A crucial geometric consequence of this principle is that no compact hypersurface can evolve smoothly forever. Indeed, any compact initial surface Σ_0 can be enclosed by a sufficiently large round sphere. Since we know that the enclosing sphere shrinks to a point in finite time, the avoidance principle forces Σ_t to remain trapped inside it. Consequently, Σ_t must vanish in finite time.

The behavior of convex surfaces under MCF is completely characterized by a foundational result in geometric analysis:

Theorem 2.1 (Huisken, 1984 [12]) *If the initial closed hypersurface Σ_0 is strictly convex, it contracts smoothly and becomes spherical under the evolution by Mean Curvature Flow, eventually collapsing to a point in finite time.*

However, if the initial hypersurface is not strictly convex, singularities may appear before the surface completely collapses to a point. A classic counterexample is the evolution of Grayson's dumbbell [10]: as the flow progresses, the central neck pinches off while the two outer lobes still have positive volume, creating a so-called “neck-pinch" singularity.

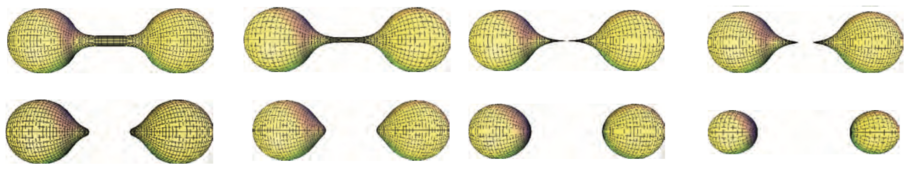


Figure 3: The evolution of Grayson's dumbbell.

2.2 Singularity analysis

As guaranteed by the maximum principle, the Mean Curvature Flow of a compact hypersurface cannot exist indefinitely. There exists a maximal finite time $T < \infty$ at which the flow develops a singularity. Analytically, the maximal time is characterized by the global blow-up of the curvature:

$$\lim_{t \nearrow T} \max_{\Sigma_t} |A|^2 = +\infty$$

To localize this phenomenon, we define a *singularity point* as a point (p, T) in spacetime $\mathbb{R}^{n+1} \times \mathbb{R}_+$ such that

$$|A_{\Sigma_t}|^2(p) \to +\infty, \quad \text{as } t \to T.$$

To understand the local geometric behavior near a singularity point (p, T), it is standard to classify the maximal flow based on the rate at which the curvature blows up. We distinguish between two categories:

- **Type I singularities** (rapidly forming, but controlled):The curvature blow-up is bounded by the natural parabolic scaling of the equation. We say the flow develops a Type I singularity if there exists a constant $C > 0$ such that:

  $$\max_{\Sigma_t} |A|^2 \leq \frac{C}{T - t}$$

  for all $t \in [0, T)$. The shrinking sphere and Grayson's neck-pinch are both examples of Type I singularities.

- **Type II singularities** (degenerate): The curvature blows up strictly faster than the Type I rate, meaning:

  $$\limsup_{t \nearrow T} \max_{\Sigma_t} |A|^2 (T - t) = +\infty$$

  An example of this kind of singularity occurs when an immersed curve develops a cusp during its evolution, leading to an extremely localized and unbounded accumulation of curvature.

In the following analysis, we focus on Type I singularities. To study what a singularity "looks like" under a microscope, one employs a parabolic rescaling procedure. By centering the coordinates at the singular point (p, T) and dilating the spatial and temporal variables, one can extract a limit flow.

The fundamental tool that ensures the convergence and the geometric rigidity of this rescaling procedure is Huisken's Monotonicity Formula [13]. It states that the integral of the backward heat kernel over the evolving surface Σ_t is strictly decreasing in time. The fundamental result is the following:

Theorem 2.2 (Huisken's Rescaling Theorem, [13]) *Let (p, T) be a Type I singularity of the Mean Curvature Flow $F(\cdot, t)$. Define the parabolic rescaling centered at (p, T) by the normalized immersions:*

$$\tilde{F}(q, s) = \frac{F(q, t) - p}{\sqrt{2(T - t)}}, \quad \textit{where} \quad s = -\frac{1}{2}\log(T - t).$$

Then, for any sequence of rescaled times $s_j \to \infty$, there exists a subsequence such that the rescaled hypersurfaces $\tilde{\Sigma}_{s_j}$ converge smoothly on compact subsets of $\mathbb{R}^{n+1}$ to a non-empty, smooth limit hypersurface Σ_∞. Moreover, this limit surface satisfies the soliton equation:

$$H = \langle x, \nu \rangle. \tag{2.1}$$

By a direct computation it's possible to show that every hypersurface Σ_∞ satisfying the soliton equation (2.1) generates a self-shrinker solution to the Mean Curvature Flow. Thus, Huisken's theorem elegantly guarantees that all Type I singularities are asymptotically modeled by self-shrinkers.

A remarkable rigidity result holds when we restrict our attention to surfaces with positive mean curvature. For the mean convex case ($H > 0$), there are fundamentally just two geometric models for Type I singularities: the generalized cylinder $\mathbb{S}^k \times \mathbb{R}^{n-k}$ and the round sphere $\mathbb{S}^n$.

For instance, in the evolution of Grayson's dumbbell discussed earlier, if we perform Huisken's rescaling procedure centered exactly at the neck-pinch singularity, the blow-up limit Σ_∞ that we extract is exactly the straight solid cylinder $\mathbb{S}^1 \times \mathbb{R}$.

3 A brief introduction to the Heisenberg group

We want to give a heuristic introduction to the Heisenberg group $\mathbb{H}^1$; for a deeper and more rigorous treatment, we refer to [5] and [1]. To understand the geometry of $\mathbb{H}^1$, imagine living in the standard three-dimensional space $\mathbb{R}^3$ with coordinates (x, y, z), but with a severe physical restriction on how you can move. Instead of being able to travel freely in any direction, you are strictly forbidden from moving straight "up" or "down" along the z-axis.

At any point $p = (x, y, z)$, your allowed directions of motion are confined to a two-dimensional plane $\mathcal{H}_p$. This plane is not flat and horizontal everywhere; instead, it "twists" as you move around the xy-plane. We can define these allowed directions by choosing two linearly independent vector fields that span $\mathcal{H}_p$:

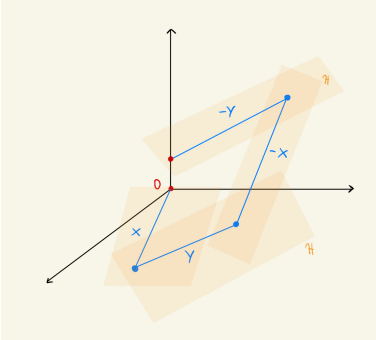
$$X = \partial_x - \frac{y}{2}\partial_z, \qquad Y = \partial_y + \frac{x}{2}\partial_z$$

The space consisting of all these planes, $\mathcal{H} = \text{span}\{X, Y\}$, is called *horizontal distribution* (specifically, a contact structure). A curve is called *horizontal* (or admissible) if its tangent vector always lies within $\mathcal{H}$.

At first glance, one might think that being confined to these twisting planes traps us in a sub-region of $\mathbb{R}^3$. However, a direct computation reveals the magic of sub-Riemannian geometry. If we compute the Lie bracket (the commutator) of the two allowed vector fields, we find:

$$[X, Y] = XY - YX = \partial_z \equiv Z$$

This algebraic relation has a profound geometric meaning: even though we cannot move directly along the Z direction, we can generate a net displacement purely along the z-axis by moving infinitesimally along X, then Y, then $-X$, and finally $-Y$.



This is exactly the mathematical formalization of parallel parking a car: a car cannot move sideways directly, but by combining forward/backward motions with steering, it can achieve a net sideways translation. Because X, Y, and their commutator $[X, Y]$ span the entire tangent space at every point, a celebrated theorem by Chow and Rashevskii (see [1, Theorem 3.31]) guarantees that any two points in $\mathbb{H}^1$ can be connected by a horizontal curve.

3.1 Hypersurfaces and Horizontal Geometry

Now that we have an intuitive grasp of the ambient space $\mathbb{H}^1$, we want to study the geometry of a smooth hypersurface $\Sigma \subset \mathbb{H}^1$. Suppose Σ is locally given as the zero level set of a smooth defining function $f : \mathbb{H}^1 \to \mathbb{R}$, so that $\Sigma = \{p \in \mathbb{H}^1 \mid f(p) = 0\}$.

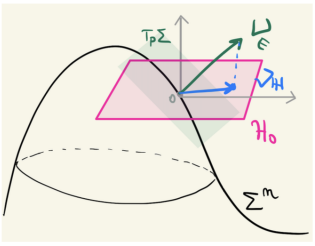
In sub-Riemannian geometry, the metric properties are determined entirely by the horizontal distribution $\mathcal{H}$. Therefore, we define the *horizontal gradient* of f as the projection of the gradient onto $\mathcal{H}$, expressed using our left-invariant vector fields:

$$\nabla_H f = (Xf)X + (Yf)Y$$

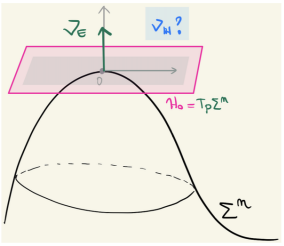
We define the *horizontal unit normal* as the projection of the Euclidean normal onto the horizontal distribution, we denote it as ν_H, and it holds that:

$$\nu_H = -\frac{\nabla_H f}{|\nabla_H f|} = \nu_1 X + \nu_2 Y$$

where $|\nabla_H f| = \sqrt{(Xf)^2 + (Yf)^2}$ is the sub-Riemannian norm.
A striking feature of sub-Riemannian geometry is the existence of *characteristic points*. These are points $p \in \Sigma$ where the horizontal plane $\mathcal{H}_p$ coincides with the classical tangent space $T_p\Sigma$. At these points, the horizontal gradient vanishes entirely ($\nabla_H f(p) = 0$) and it is not possible to define the horizontal unit normal, even though the surface is smooth in the Euclidean sense. The set of all such points is called the characteristic set $\mathrm{Char}(\Sigma)$.



(A) Horizontal unit normal



(B) Characteristic point

Just as the classical mean curvature is the divergence of the Euclidean normal, the *horizontal mean curvature* H_H is defined as the horizontal divergence of the horizontal unit normal:

$$H_H = \mathrm{div}_H(\nu_H) = X(\nu_1) + Y(\nu_2)$$

This quantity H_H governs the geometric analysis of surfaces in $\mathbb{H}^1$. It represents the trace of the symmetrized horizontal second fundamental form and, crucially, it appears as the first variation of the sub-Riemannian perimeter (see [5, Proposition 6.24]).

4 Horizontal Mean Curvature Flow

Since ν_H, and consequently H_H, are not well defined at every point of a regular hypersurface, we need a weaker definition of the Mean Curvature Flow in $\mathbb{H}^1$, the so-called *Horizontal Mean Curvature Flow* (HMCF).
Classically, for the Euclidean Mean Curvature Flow, there are two ways to extend the flow through singularities by defining a weak flow. The first one is due to Brakke [3] and exploits objects from geometric measure theory, such as varifolds. The second approach is the one we will adapt in our sub-Riemannian context; it appeared originally in two independent works by Evans and Spruck [7], and Chen, Giga, and Goto [6]. This second approach is the *level set method* and the first work that adapted this approach to the sub-Riemannian setting is due to Capogna and Citti [4]. In the following analysis, we will focus on the level set approach for strictly horizontal mean convex hypersurfaces, i.e., surfaces where $H_H \geq c > 0$ outside characteristic points.

The key idea is to represent the evolving hypersurface Σ_t as the t-level set of a continuous scalar function $u : \bar{\Omega} \to \mathbb{R}$, where $\Omega \subset \mathbb{H}^1$ is the region enclosed by the initial surface.

Specifically, we define:

$$\Sigma_t = \{p \in \bar{\Omega} \mid u(p) = t\}.$$

In this context, u is known as the *arrival time function*, as $u(p)$ represents the exact time the shrinking surface passes through the point p.

By requiring that every level set of u evolves according to the Horizontal Mean Curvature Flow with velocity equal to $-\nu_H H_H$, and starting from a strictly horizontal mean convex initial hypersurface $\Sigma_0 = \partial\Omega$, one can derive a scalar boundary value problem for u. The function u must satisfy the following:

$$\begin{cases} -1 = |\nabla_H u| \operatorname{div}_H \left(\frac{\nabla_H u}{|\nabla_H u|} \right) & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega. \end{cases} \tag{4.1}$$

This is a highly degenerate, sub-elliptic equation. The degeneracy occurs at all characteristic points, where the horizontal gradient vanishes ($\nabla_H u = 0$). If u is a solution to (4.1), then $\Sigma_t = \{p \in \mathbb{H}^1 \mid u(p) = t\}$ evolves by the HMCF.

To rigorously make sense of solutions to (4.1) across these singular sets, one must employ the theory of *viscosity solutions*. The definition we adopt is the natural analogue to the one presented in [4]; for the sake of brevity, we will not go into further technical details here.

Remark 4.1 We want to emphasize that equation (4.1) makes sense only under the strict horizontal mean convexity assumption. Indeed, along the flow, the equation can be rewritten to explicitly show that the horizontal mean curvature is strictly positive:

$$H_H = \frac{1}{|\nabla_H u|} > 0.$$

Now we state our existence result for such problem.

Theorem 4.2 *Let $\Omega \subset \mathbb{H}^1$ be a bounded, C^2-smooth, strictly horizontal mean convex set, such that the horizontal mean curvature $H_H(\partial\Omega)$ extends continuously to the characteristic points. Assume that the boundary $\partial\Omega$ satisfies one of the following conditions:*

- $\mathrm{Char}(\partial\Omega) = \emptyset$,
- *Axisymmetric (invariant under rotations around the z-axis).*

Then, there exists a viscosity solution to (4.1) that is Lipschitz in a suitable sense.

4.1 Self-shrinkers for HMCF

In parallel to the Euclidean setting, we want to define self-shrinkers for the Horizontal Mean Curvature Flow. However, in the sub-Riemannian context of the Heisenberg group, dilations must respect the Lie algebra stratification. Therefore, we introduce the anisotropic Heisenberg dilations, defined for any $\lambda > 0$ as:

$$\delta_\lambda(x, y, z) = (\lambda x, \lambda y, \lambda^2 z). \tag{0.1}$$

A hypersurface Σ is called a *self-shrinker* for the HMCF if its evolution is given purely by these anisotropic dilations, namely $\Sigma_t = \delta_{\lambda(t)}\Sigma$.

To obtain a non-existence theorem for self-shrinkers of HMCF we study the behaviour of the flow at characteristic points, in particular we obtain the following:

Proposition 4.3 *Let u be a solution to (4.1) with initial datum $\Omega \subset \mathbb{H}$ which is smooth, open, bounded and $\{\Gamma_t\} = \{x \in \Omega \mid u(x) = t\}$ is a C^2 foliation. Then, for any $t > 0$, the horizontal mean curvature of Γ_t satisfies*

$$H_H(p) \to \infty,$$

when p approaches a characteristic point of Γ_t.

As a consequence we obtain a non-existence result.

Corollary 4.1 *Let u be a solution to (4.1) with initial datum $\Omega \subset \mathbb{H}$ which is smooth, open, bounded. If $\partial\Omega$ has at least one characteristic point and bounded horizontal mean curvature $H_{\mathbb{H}} < c < \infty$. Then, Ω is not a self-shrinker for HMCF.*

This result highlights a break point from the Euclidean case, and its most striking application concerns the *Pansu spheres*.

In the Heisenberg group, the Pansu spheres are the conjectured isoperimetric sets, playing the same role as round spheres in $\mathbb{R}^3$. However, a Pansu sphere has two characteristic points (its "poles") and constant horizontal mean curvature equal to $\frac{1}{R}$. By our corollary, the horizontal mean curvature blows up at these poles, meaning the Pansu sphere *cannot* be a self-shrinker.

Unlike the Euclidean sphere, which shrinks to a point homothetically, the Pansu sphere inevitably distorts as it evolves under HMCF.

4.2 Soliton equation and a conjecture

Finally, we were able to write a counterpart to Huisken's soliton equation (2.1) in the Heisenberg group.

Theorem 4.5 *Let $M_0 = \{\phi = 0\} \subset \mathbb{H}^1$ be a smooth hypersurface. Assume that the rescaled family $\delta_{\lambda(t)}M_0$, with $\lambda(t) = \sqrt{1-2t}$, evolves by HMCF. Then, outside the characteristic set, the level set function ϕ satisfies the soliton equation:*

$$D^{\perp}\phi = H_H(M_0)\,|\nabla_H \phi|$$

where $D^{\perp}$ is the Euclidean orthogonal projection of the dilation vector field.

The only exact solution that, for the moment, we were able to find for this equation is the vertical cylinder, which acts as a self-shrinker for the HMCF.

This strongly motivates our conjecture about the asymptotic behavior of the flow. To build the intuition behind it, we focus on the explicit evolution of the Koranyi ball, studied by Ferrari, Liu, and Manfredi [8]. The Koranyi ball is a regular hypersurface defined as the zero level set of the following homogeneous norm:

$$\Sigma_0 = \{(x_1^2 + x_2^2)^2 + 16x_3^2 = R^4\}.$$

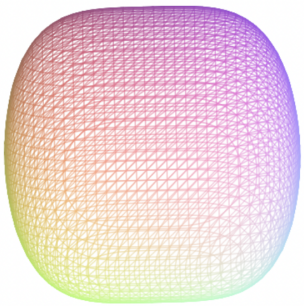


Figure 5: The Koranyi ball in the Heisenberg group.

Initially, at the characteristic points ($x_1 = x_2 = 0$), the horizontal mean curvature vanishes: $H_H(\Sigma_0) \to 0$. However, as soon as the flow starts ($t > 0$), the curvature instantaneously explodes at the poles:

$$H_H(\Sigma_t) = \frac{12(x_1^2 + x_2^2) + 24t}{|\nabla_H u_t|} \longrightarrow \infty.$$

As t approaches the extinction time T, the surface shrinks completely. If we perform an anisotropic rescaling of the surface by a factor $\delta_{1/\lambda(t)^2}$, where $\lambda(t) = \sqrt{R^4 - 12t^2} \to 0$, we can extract the geometric limit. The rescaled surfaces converge exactly to a straight vertical cylinder:

$$\delta_{1/\lambda(t)^2}(\Sigma_t) \longrightarrow \left\{ \xi_1^2 + \xi_2^2 = \frac{1}{R^2\sqrt{12}} \right\} \quad \text{as } t \to T.$$

The fact that the Koranyi ball becomes asymptotically a cylinder, combined with the behavior at characteristic points, leads to the following fundamental open question:

Conjecture: The evolution by Horizontal Mean Curvature Flow of any generic mean convex initial hypersurface tends asymptotically to a cylinder at the singular time.

References

[1] A. Agrachev, D. Barilari, and U. Boscain, "A Comprehensive Introduction to Sub-Riemannian Geometry". Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2019.

[2] G. Bombardieri, M. Fogagnolo, and V. Franceschi, *Mean convex mean curvature flow in the Heisenberg group.* In preparation, 2026.

[3] K.A. Brakke, "The Motion of a Surface by Its Mean Curvature. (MN-20)". Princeton University Press, 1978.

[4] L. Capogna and G. Citti, *Generalized mean curvature flow in Carnot groups.* Communications in Partial Differential Equations, 34(8): 937–956, 2009.

[5] L. Capogna, D. Danielli, S. Pauls, and J. Tyson, "An Introduction to the Heisenberg Group and the Sub-Riemannian Isoperimetric Problem". Progress in Mathematics. Birkhäuser Basel, 2007.

[6] Y.G. Chen, Y. Giga, and S. Goto, *Uniqueness and existence of viscosity solutions of generalized mean curvature flow equations.* Journal of Differential Geometry, 33(3): 749–786, 1991.

[7] L.C. Evans and J. Spruck, *Motion of level sets by mean curvature.* I. Journal of Differential Geometry, 33(3): 635–681, 1991.

[8] F. Ferrari, Q. Liu, and J.J. Manfredi, *On the horizontal mean curvature flow for axisymmetric surfaces in the Heisenberg group.* Communications in Contemporary Mathematics, 16(03): 1350027, May 2014.

[9] M. Gage and R.S. Hamilton, *The heat equation shrinking convex plane curves.* Journal of Differential Geometry, 23(1): 69–96, 1986.

[10] M.A. Grayson, *A short note on the evolution of a surface by its mean curvature.* Duke Mathematical Journal, 58(3): 555–558, 1989.

[11] M.A. Grayson, *Shortening embedded curves.* Annals of Mathematics, 129(1) :71–111, 1989.

[12] G. Huisken, *Flow by mean curvature of convex surfaces into spheres.* Journal of Differential Geometry, 20(1): 237–266, 1984.

[13] G. Huisken, *Asymptotic behavior for singularities of the mean curvature flow.* Journal of Differential Geometry, 31(1): 285–299, 1990.

[14] C. Mantegazza, "Lecture Notes on Mean Curvature Flow". Volume 290 of Progress in Mathematics. Birkhäuser Basel, 2011.

Effect of bond disorder on the supercritical discrete Gaussian free field

Emanuele Pasqui (*)

Abstract. The discrete Gaussian free field is a model of random variables on graphs that is central in statistical mechanics, and can be interpreted as a microscopic description of fluctuations in a homogeneous crystal at non-zero temperature. We define the model on $\mathbb{Z}^d$ when $d \geq 3$, and discuss the so-called hard wall event, in which the field is constrained to stay non-negative on a given set. We study both the decay of its probability and the behaviour of the field conditioned on this event, as the size of the set grows to infinity. Finally, we introduce disorder into the model in the form of random weights on the edges of the graph, modelling inhomogeneities in the crystal, and analyse how this disorder affects the aforementioned results.

1 Introduction

Consider an undirected weighted graph $G = (V_G, E_G, \omega_G)$, where V_G is a set of vertices, E_G is a set of edges between the vertices in V_G and $\omega_G \in [0, \infty)^{E_G}$ is a map assigning a non-negative weight to each edge in E_G. The Gaussian free field on G is a centered Gaussian process $\varphi = (\varphi_x)_{x \in V_G}$ with covariance matrix given by the Green's function of a random walk on G which has probability of crossing any given edge $\{x, y\} \in E_G$ that is proportional to the weight $\omega_{x,y}$ on that edge.

We analyse how *impurities* affect the behaviour of the discrete Gaussian free field on $\mathbb{Z}^d$, $d \geq 3$, in the presence of a hard wall constraint. This event corresponds to the atypical scenario in which the field stays non-negative over a macroscopic domain. Given the classic, homogeneous model where all the edges of the graph have the same weight, we introduce an inhomogeneous model where weights are not assumed to be equal, and they are randomised according to a probability law $\mathbb{Q}$ which is stationary and ergodic with respect to spatial translations on $\mathbb{Z}^d$, and that has enough α-mixing properties. For the randomised setting, we are interested in the *quenched* behaviour of the field, that is, those properties that hold for $\mathbb{Q}$-almost all realisations of the random weights.

(*)Ph.D. course, Università di Padova, Dip. Matematica, via Trieste 63, I-35121 Padova, Italy. E-mail: pasqui@math.unipd.it . Seminar held on 29 January 2026.

Let $V \subseteq \mathbb{R}^d$ be a compact set, and for $N \in \mathbb{N}$ define its discrete blow-up as

$$V_N = (NV) \cap \mathbb{Z}^d \,. \tag{1.1}$$

We study the hard wall event, given by

$$\mathcal{W}_N^+ = \{\varphi_x \geq 0 \,,\, \forall x \in V_N\} \,, \tag{1.2}$$

that is the event that the field stays above a hard wall at height 0 on V_N. We focus on the decay of the probability of $\mathcal{W}_N^+$ as $N \to \infty$ and on the behaviour of the field conditioned on $\mathcal{W}_N^+$. For the latter, we first analyse the asymptotic profile of the expectation of the conditioned field, and then we provide a pathwise characterisation for the conditioned law (see (3.1), (3.3) and (3.4) for these results).

We now describe the organisation of these notes. In Section 2 we introduce all the necessary preliminaries. In Section 3 the main results are stated, and the effect of disorder is discussed. Finally, Section 4 contains a sketch of the proofs for the decay of the hard wall probability. In particular, for the lower bound, a naive idea of proof is presented, which provides a non-sharp lower bound, together with the reason why it is not sharp. For the upper bound, a sketch of proof for the sharp result is presented exploiting a spatial Markov property of the Gaussian free field.

2 Notation and useful results

2.1 Notation

We fix the dimension as an integer $d \geq 3$ and denote by $|\cdot|$ and by $|\cdot|_\infty$ respectively the Euclidean and ℓ^∞-distance on $\mathbb{R}^d$. For $K \subseteq \mathbb{Z}^d$, we denote by $|K|$ the cardinality of K.

We consider the integer lattice, that is the graph with vertex set $\mathbb{Z}^d$ and edge set $E_d = \{\{x, y\} \in \mathbb{Z}^d \times \mathbb{Z}^d \,:\, |x - y| = 1\}$. We write $x \sim y$ if $\{x, y\} \in E_d$, and say in this case that x and y are neighbours. We also consider two real numbers $0 < \lambda \leq \Lambda$ and the set $\Omega_{\lambda,\Lambda} = [\lambda, \Lambda]^{E_d}$ of weight configurations on the edges uniformly bounded away from zero and infinity. We will also refer to such weights as conductances, and to $\omega \in \Omega_{\lambda,\Lambda}$ as an environment.

2.2 Elements of potential theory

Given an environment $\omega \in \Omega_{\lambda,\Lambda}$, we introduce the measure

$$\mu_x^\omega = \sum_{y \sim x} \omega_{x,y} \,, \qquad x \in \mathbb{Z}^d \,, \tag{2.1}$$

and the continuous-time constant-speed random walk on the weighted graph $(\mathbb{Z}^d, E_d, \omega)$ as the Markov process $X = (X_t)_{t \geq 0}$ with generator $\mathcal{L}^\omega : \mathbb{R}^{\mathbb{Z}^d} \to \mathbb{R}^{\mathbb{Z}^d}$ defined by

$$\mathcal{L}^\omega f(x) = \sum_{y \sim x} \frac{\omega_{x,y}}{\mu_x^\omega} (f(y) - f(x)) \,, \qquad f \in \mathbb{R}^{\mathbb{Z}^d} \,,\, x \in \mathbb{Z}^d \,. \tag{2.2}$$

For $x \in \mathbb{Z}^d$, we denote by P_x^ω the law of X started at x, and by E_x^ω the associated expectation. Under P_x^ω, the process X is a Markov chain such that $X_0 = x$, and at time t it waits on the vertex X_t for an exponential time with parameter 1 and then jumps to a neighbour y of X_t with probability $\omega_{X_t,y}/\mu^\omega_{X_t}$.

For $U \subseteq \mathbb{Z}^d$, we define the stopping times (with respect to the canonical filtration generated by X) $H_U = \inf\{t \geq 0 : X_t \in U\}$ and $T_U = \inf\{t \geq 0 : X_t \notin U\}$, which are the entrance and exit times of U, respectively. We define the Green function and the killed Green function upon exiting U associated to P_x^ω, namely $g^\omega, g_U^\omega : \mathbb{Z}^d \times \mathbb{Z}^d \to [0,\infty)$, respectively, as

(2.3)
$$g^\omega(x,y) = E_x^\omega\Big[\int_0^\infty \mathbb{1}_{\{X_t=y\}}\mathrm{d}t\Big]/\mu_y^\omega \qquad \text{and} \qquad g_U^\omega(x,y) = E_x^\omega\Big[\int_0^{T_U} \mathbb{1}_{\{X_t=y\}}\mathrm{d}t\Big]/\mu_y^\omega\,,$$

for $x, y \in \mathbb{Z}^d$. It is well-known that the walk is transient for $d \geq 3$, and $g(x,x) < \infty$ for all $x \in \mathbb{Z}^d$ (and this also holds for $g(x,y)$ for all $x, y \in \mathbb{Z}^d$, since $g(x,\cdot)$ is harmonic outside x). Moreover, the following bounds hold.

$$\frac{c_1}{|x-y|^{d-2}\vee 1} \leq g^\omega(x,y) \leq \frac{c_2}{|x-y|^{d-2}\vee 1}\,, \qquad x, y \in \mathbb{Z}^d\,,\, \omega \in \Omega_{\lambda,\Lambda}\,, \tag{2.4}$$

where the constants c_1, c_2 do not depend on the choice of $\omega \in \Omega_{\lambda,\Lambda}$, see Theorem 6.28 of [1]. We now define the capacity of a bounded set $A \subseteq \mathbb{Z}^d$ for the random walk X as

$$\mathrm{Cap}^\omega(A) = \lim_{|x|\to\infty} \frac{P_x^\omega[H_A < \infty]}{g^\omega(x,0)}\,, \tag{2.5}$$

see e.g. page 59 of [9]. We further introduce the equilibrium potential for A associated to X as the function

$$h_A^\omega(x) = P_x^\omega[H_A < \infty]\,, \qquad x \in \mathbb{Z}^d\,, \tag{2.6}$$

which is ω-harmonic on $\mathbb{Z}^d \setminus A$ and takes value 1 on A.

2.3 Homogenisation

We consider the σ-algebra of cylinders $\mathcal{G}$ for $\Omega_{\lambda,\Lambda}$ and define the group of shifts for weight configurations, $\tau_x : \Omega_{\lambda,\Lambda} \to \Omega_{\lambda,\Lambda}$, $x \in \mathbb{Z}^d$, as

$$(\tau_x\omega)_{y,z} = \omega_{y+x,z+x}\,, \qquad y, z \in \mathbb{Z}^d\,,\, \omega \in \Omega_{\lambda,\Lambda}\,. \tag{2.7}$$

We also let $\mathbb{Q}$ be a probability measure on $(\Omega_{\lambda,\Lambda}, \mathcal{G})$ that we assume to be stationary and ergodic with respect to $(\tau_x)_{x\in\mathbb{Z}^d}$, namely, respectively, that $\mathbb{Q}[\tau_x(A)] = \mathbb{Q}[A]$ for all $A \in \mathcal{G}$ and $x \in \mathbb{Z}^d$, and that $\mathbb{Q}[A] \in \{0,1\}$ for all $A \in \mathcal{G}$ with $\tau_x(A) = A$ for every $x \in \mathbb{Z}^d$.

It is then known that for $\mathbb{Q}$-a.e. $\omega \in \Omega_{\lambda,\Lambda}$, the diffusively rescaled random walk $(N^{-1}X_{N^2t})_{t\geq 0}$ under P_0^ω converges in law to a Brownian motion $Z = (Z_t)_{t\geq 0}$ with non-degenerate covariance matrix $a_\mathbb{Q} \in \mathbb{R}^{d\times d}$. Denoting by W_x the law of Z started at $x \in \mathbb{R}^d$ and considering its Green function defined as

$$(x,y) = c_{d,a_\mathbb{Q}}[(y-x)^\top a_\mathbb{Q}^{-1}(y-x)]^{(2-d)/2}\,, \qquad x, y \in \mathbb{R}^d\,,\, x \neq y\,, \tag{2.8}$$

where $c_{d,a_{\mathbb{Q}}}$ is a positive fixed constant, we define the capacity of a compact set $K \subseteq \mathbb{R}^d$ for Z as

$$\mathrm{Cap}^{a_{\mathbb{Q}}}(K) = \lim_{|x|\to\infty} \frac{W_x[H_K^Z < \infty]}{(x,0)}\,, \tag{2.9}$$

where $H_K^Z = \inf\{t \geq 0 : Z_t \in K\}$ is the entrance time of K for Z. We will refer to $\mathrm{Cap}^{a_{\mathbb{Q}}}(K)$ as the homogenised capacity of K under the law $\mathbb{Q}$. We further introduce the equilibrium potential for K associated to Z as the function

$$\overset{a_{\mathbb{Q}}}{K}(x) = W_x[H_K^Z < \infty]\,, \qquad x \in \mathbb{R}^d\,, \tag{2.10}$$

which is harmonic on $\mathbb{R}^d \setminus K$ and takes value 1 on K. Then, a homogenisation result holds, that is, for every compact $A \subseteq \mathbb{R}^d$, for $\mathbb{Q}$-a.a. $\omega \in \Omega_{\lambda,\Lambda}$

$$\lim_{N\to\infty} \frac{\mathrm{Cap}^{\omega}(A_N)}{N^{d-2}} = \mathrm{Cap}^{a_{\mathbb{Q}}}(A) \tag{2.11}$$

(see Proposition 5.3 and Corollary 5.4 of [5], whose proofs are based on the argument of Corollary 3.4 of [8].

2.4 The discrete Gaussian free field

Given an environment $\omega \in \Omega_{\lambda,\Lambda}$, we define the discrete Gaussian free field on $(\mathbb{Z}^d, E_d, \omega)$ as the centered Gaussian field $\varphi = (\varphi_x)_{x\in\mathbb{Z}^d}$ with covariance given by the Green function g^ω, that is, denoting by $\mathbb{P}^\omega$ its law and by $\mathbb{E}^\omega$ the associated expectation,

$$\mathbb{E}^\omega[\varphi_x] = 0 \quad \text{and} \quad \mathbb{E}^\omega[\varphi_x\varphi_y] = g^\omega(x,y)\,, \tag{2.12}$$

for all $x, y \in \mathbb{Z}^d$.

We now introduce a spatial Markov property for the Gaussian free field. For $U \subseteq \mathbb{Z}^d$, we define the (ω-)harmonic average $\xi^{\omega,U}$ of φ in U as

$$\xi_x^{\omega,U} = E_x^\omega[\varphi_{X_{T_U}}, T_U < \infty] = \sum_{y\in\mathbb{Z}^d} P_x^\omega[X_{T_U} = y, T_U < \infty]\varphi_y\,, \qquad x \in \mathbb{Z}^d\,, \tag{2.13}$$

and the (ω-)local field $\psi^{\omega,U}$ as

$$\psi^{\omega,U} = \varphi - \xi^{\omega,U}\,. \tag{2.14}$$

Note that for $x \in U^c$ we have $\psi_x^{\omega,U} = 0$ and $\xi_x^{\omega,U} = \varphi_x$. The Markov property is the following:

$$\begin{gathered}(\psi_x^{\omega,U})_{x\in\mathbb{Z}^d} \text{ is independent of } \mathcal{F}_{U^c} \text{ (in particular of } \xi^{\omega,U}),\\ \text{and is distributed as a centered Gaussian field with covariance } g_U^\omega(\cdot,\cdot)\,,\end{gathered} \tag{2.15}$$

see Proposition 2.3 of [10].

3 Main results

We assume that the conductances are randomised according to a probability law $\mathbb{Q}$ as in Subsection 2.3, namely that $\mathbb{Q}$ is stationary and ergodic with respect to translations on $\mathbb{Z}^d$. The following large deviation result was shown in [6]. If $\mathbb{Q}$ satisfies an α-mixing condition[(1)] for some $\alpha > d(d-1)/2$, then for $\mathbb{Q}$-a.e. $\omega \in \Omega_{\lambda,\Lambda}$

$$\lim_{N\to\infty} \frac{1}{N^{d-2}\log N} \log \mathbb{P}^\omega[\mathcal{W}_N^+] = -2\overline{g}_{\mathbb{Q}} \mathrm{Cap}^{a_{\mathbb{Q}}}(V)\,, \tag{3.1}$$

with $\overline{g}_{\mathbb{Q}}$ defined as

$$\overline{g}_{\mathbb{Q}} = \sup_{x\in\mathbb{Z}^d} g^\omega(x,x) = \operatorname*{ess\,sup}_{\omega\in\Omega_{\lambda,\Lambda}} g^\omega(0,0) < \infty \tag{3.2}$$

where the last equality follows from stationarity and ergodicity of $\mathbb{Q}$, which imply that, for all $x \in \mathbb{Z}^d$, $g^\omega(x,x)$ is constant for $\mathbb{Q}$-a.e. $\omega \in \Omega_{\lambda,\Lambda}$. In the same work, two results were obtained on the field conditioned on the hard wall event. The first concerns the first order behaviour of the expectation of the field conditioned on the hard wall event, as

$$\lim_{N\to\infty} \sup_{x\in\mathbb{Z}^d} \left| \frac{\mathbb{E}^\omega[\varphi_x|\mathcal{W}_N^+]}{\sqrt{4\overline{g}_{\mathbb{Q}}\log N}} - \overset{a_{\mathbb{Q}}}{V}\Big(\frac{x}{N}\Big) \right| = 0 \tag{3.3}$$

for $\mathbb{Q}$-a.a. $\omega \in \Omega_{\lambda,\Lambda}$. The second result is a pathwise characterisation of the conditioned field. Defining, for $f \in \mathbb{R}^{\mathbb{Z}^d}$, the shift operator $T_f : \mathbb{R}^{\mathbb{Z}^d} \to \mathbb{R}^{\mathbb{Z}^d}$ as $T_f\varphi = \varphi + f$, it holds that for $\mathbb{Q}$-a.e. $\omega \in \Omega_{\lambda,\Lambda}$

$$\lim_{N\to\infty} (T_{a_N})_\sharp \mathbb{P}^\omega[\,\cdot\,|\mathcal{W}_N^+] = \mathbb{P}^\omega \tag{3.4}$$

in the sense of weak convergence of measures, where a_N has the same first order behaviour as $\mathbb{E}^\omega[\varphi_x|\mathcal{W}_N^+]$ for $x \in V_N$, hence by (3.3)

$$\lim_{N\to\infty} \frac{a_N}{\sqrt{4\overline{g}_{\mathbb{Q}}\log N}} = 1\,. \tag{3.5}$$

While (3.1) quantifies the decay of the hard wall probability, (3.3) implies that the field, when conditioned on $\mathcal{W}_N^+$, undergoes an entropic repulsion away from the zero height. The first order behaviour of the repulsion is $\sqrt{4\overline{g}_{\mathbb{Q}}\log N}$ on V_N, and decays like $|x|^{-2}$ for $x \notin V_N$, as $|x|\to\infty$. (3.4) implies that the conditioned field, when recentered, converges to the Gaussian free field $\mathbb{P}^\omega$, while (3.5) yields that this recentering can be made constant across all vertices of $\mathbb{Z}^d$. This is because the difference $\mathbb{E}^\omega[\varphi_x - \varphi_y|\mathcal{W}_N^+]$ goes to 0 as $N\to\infty$, for every $x, y \in \mathbb{Z}^d$.

In the homogeneous setting, that is, when $\lambda = \Lambda$, so that the conductances are all equal, we will omit for simplicity the superscripts ω and $a_{\mathbb{Q}}$ for all the quantities defined

[(1)] $\mathbb{Q}$ satisfies an α-mixing property for $\alpha \geq 0$ if $|\mathbb{Q}[A\cap B] - \mathbb{Q}[A]\mathbb{Q}[B]| \leq d_\infty(U,V)^{-\alpha}$ for all disjoint Borel sets $U, V \subseteq \mathbb{Z}^d$ and every $A \in \sigma(\omega_{x,y} : x, y \in U)$ and $B \in \sigma(\omega_{x,y} : x, y \in V)$.

in Section 2 and write $(\mathbb{Z}^d, E_d)$ in place of $(\mathbb{Z}^d, E_d, \omega)$. It is straightforward to see that in the homogeneous setting $\mu_x = 2d$ for all $x \in \mathbb{Z}^d$ and $P_{x_1}[X_1 = x_2] = \mathbb{1}_{\{x_1 \sim x_2\}}/2d$ for every $x_1, x_2 \in \mathbb{Z}^d$, hence g is translation invariant, that is

$$g(x, y) = g(x - y, 0)\,, \qquad x, y \in \mathbb{Z}^d\,. \tag{3.6}$$

Consequently, in this scenario the field has the same variance at every site. In this setting, (3.1) and (3.3) were shown in [3], respectively as

$$\lim_{N\to\infty} \frac{1}{N^{d-2}\log N} \log \mathbb{P}[\mathcal{W}_N^+] = -2g(0,0)\mathrm{Cap}(V) \tag{3.7}$$

and

$$\lim_{N\to\infty} \sup_{x\in V_N} \left| \frac{\mathbb{E}[\varphi_x | \mathcal{W}_N^+]}{\sqrt{4g(0,0)\log N}} - 1 \right| = 0\,, \tag{3.8}$$

while (3.4) and (3.5) were shown in [7] as

$$\lim_{N\to\infty} (T_{a_N})_\sharp \mathbb{P}[\,\cdot\,|\mathcal{W}_N^+] = \mathbb{P}\,, \qquad \lim_{N\to\infty} \frac{a_N}{\sqrt{4g(0,0)\log N}} = 1\,. \tag{3.9}$$

We now explain the effect of disorder on the previous results. Note that the rate of decay of the hard wall probability in (3.1) is governed by two quantities: the homogenised capacity of V (defined in (2.9)), concerning the homogenised behaviour of the Brownian motion Z to which the diffusively rescaled walk X converges, for $\mathbb{Q}$-a.e. environment $\omega \in \Omega_{\lambda,\Lambda}$, and the constant $\overline{g}_{\mathbb{Q}}$ (defined in (3.2)), concerning the behaviour of the random walk X in the environment $\omega \in \Omega_{\lambda,\Lambda}$ prior to homogenisation. Hence, $\mathrm{Cap}^{a_{\mathbb{Q}}}(V)$ depends on $a_{\mathbb{Q}}$, while $\overline{g}_{\mathbb{Q}}$ depends on $\mathbb{Q}$ itself. This implies that, changing the law $\mathbb{Q}$ of the conductances, one can change $\overline{g}_{\mathbb{Q}}$ without changing $\mathrm{Cap}^{a_{\mathbb{Q}}}(V)$. In other words, one can find $\mathbb{Q}_1$ and $\mathbb{Q}_2$ such that $a_{\mathbb{Q}_1} = a_{\mathbb{Q}_2}$, hence $\mathrm{Cap}^{a_{\mathbb{Q}_1}}(V) = \mathrm{Cap}^{a_{\mathbb{Q}_2}}(V)$, but $\overline{g}_{\mathbb{Q}_1} \neq \overline{g}_{\mathbb{Q}_2}$, so that the rate of decay of the hard wall probability in (3.1) is different for $\mathbb{Q}_1$ and $\mathbb{Q}_2$. The same happens for the first order behaviour of the entropic push in (3.3), which is governed by both $\overline{g}_{\mathbb{Q}}$, concerning X, and $\overset{a_{\mathbb{Q}}}{V}$, concerning Z. In the homogeneous setting, this is obviously not the case.

In particular, consider the case where $\mathbb{Q}$ is such that conductances are i.i.d. and the essential infimum of their support is λ. Then it can be shown that $\overline{g}_{\mathbb{Q}} = g^1(0,0)/\lambda$ where g^1 is the Green function for the walk when conductances are all equal to 1 (see (6.25) in [6]). Since $a_{\mathbb{Q}}$ is a non-degenerate covariance matrix, we can write $a_{\mathbb{Q}} = \sigma^2\mathrm{Id}$ with $\sigma^2 > 0$. Considering now the homogeneous environment with $\omega(\sigma)_e = \sigma^2/2$ for all $e \in E_d$, we have that the covariance matrix of the limiting Brownian motion for the diffusively rescaled walk on $(\mathbb{Z}^d, E_d, \omega(\sigma))$ is still $\sigma^2\mathrm{Id}$. However, if the conductances are not deterministic under $\mathbb{Q}$ then denoting by $\mathbb{E}^{\mathbb{Q}}$ the expectation associated to $\mathbb{Q}$, for any given $x \sim 0$ we have

$$\lambda < \frac{1}{\mathbb{E}^{\mathbb{Q}}[\omega_{0,x}^{-1}]} \leq \frac{\sigma^2}{2} \leq \mathbb{E}^{\mathbb{Q}}[\omega_{0,x}] < \Lambda \tag{3.10}$$

(see for example [2, Section 4]), hence we get $\overline{g}_{\mathbb{Q}} = g^1(0,0)/\lambda > 2g^1(0,0)/\sigma^2 = g^{\omega(\sigma)}(0,0)$. This shows that disorder increases the maximal variance, hence modifies the rate of decay of the hard wall probability and the first order behaviour of the entropic push for the conditioned field, even if the homogenised capacity of V is the same as in the homogeneous environment $\omega(\sigma)$.

4 Sketch of proof for the hard wall probability decay

In this section we present the main ideas behind the proof of (3.1). The argument consists of establishing matching lower and upper bounds.

4.1 A non-sharp lower bound

To prove the lower bound in (3.1), one uses a shift of measure approach, that is, considering a probability measure $\widetilde{\mathbb{P}}$ absolutely continuous with respect to $\mathbb{P}^\omega$ and such that $\lim_{N\to\infty} \widetilde{\mathbb{P}}[\mathcal{W}_N^+] = 1$, and making use of the relative entropy inequality

$$\log \frac{\mathbb{P}^\omega[F]}{\widetilde{\mathbb{P}}[F]} \geq -\frac{1}{\widetilde{\mathbb{P}}[F]}\Big[H(\widetilde{\mathbb{P}}|\mathbb{P}^\omega) + \mathrm{e}^{-1}\Big]\,, \tag{4.1}$$

for any event F with positive $\widetilde{\mathbb{P}}$-probability, to get a lower bound on $\mathbb{P}^\omega[\mathcal{W}_N^+]$, where in (4.1), denoting by $\widetilde{\mathbb{E}}$ the expectation associated to $\widetilde{\mathbb{P}}$, $H(\widetilde{\mathbb{P}}|\mathbb{P}^\omega) = \widetilde{\mathbb{E}}[\log(\mathrm{d}\widetilde{\mathbb{P}}/\mathrm{d}\mathbb{P}^\omega)] \in [0,\infty)$ is the relative entropy of $\widetilde{\mathbb{P}}$ with respect to $\mathbb{P}^\omega$.

In order to get such a measure $\widetilde{\mathbb{P}}$, one could think of lifting the expectation of the field entries over V_N by the first order behaviour of $\mathbb{E}^\omega[\min_{x\in V_N}\varphi_x] = -\mathbb{E}^\omega[\max_{x\in V_N}\varphi_x]$ (the equality holding since the field is centered), so that the first order of the expected maximal fluctuation over V_N of the resulting field stays non-negative. In this way the analysis of the extremes of the field becomes relevant in the study of the hard wall event. Such first order behaviour is given by the following limit, that holds for $\mathbb{Q}$-a.e. $\omega \in \Omega_{\lambda,\Lambda}$.

$$\lim_{N\to\infty} \frac{\mathbb{E}^\omega[\max_{x\in V_N}\varphi_x]}{\sqrt{2d\overline{g}_{\mathbb{Q}}\log N}} = 1\,, \tag{4.2}$$

see (4.30) in [6]. In the homogeneous case, the scaling limit for the random variable $\max_{x\in V_N}\varphi_x$ was found in [4, Theorem 1.1] as

$$\lim_{N\to\infty} \mathbb{P}\Big[\frac{\max_{x\in V_N}\varphi_x - b_N}{a_N} < z\Big] = \exp(-\mathrm{e}^{-z})\,, \qquad z\in\mathbb{R}\,, \tag{4.3}$$

where

$$b_N = \sqrt{g(0,0)}\Big[\sqrt{2d\log N} - \frac{\log(d\log N) + \log(4\pi)}{2\sqrt{2d\log N}}\Big] \quad \text{and} \quad a_N = \frac{g(0,0)}{b_N}\,, \tag{4.4}$$

that is, the fluctuations of the maximum after recentering and scaling converge to a Gumbel distribution.

Thus, we want to shift the Gaussian free field by a field $\phi_N = f_N\sqrt{2d\overline{g}_{\mathbb{Q}}\log N}$, where $f_N \in \mathbb{R}^{\mathbb{Z}^d}$ minimises the cost of shifting, given by the absolute value of the numerator in the right-hand side of (4.1), among all the functions in $\mathbb{R}^{\mathbb{Z}^d}$ that take value 1 over V_N. Such cost is minimised when $H(\widetilde{\mathbb{P}}|\mathbb{P}^\omega)$ is minimised. Since the Radon-Nikodym derivative of the shifted Gaussian free field is known (see (2.3) and (2.4) in [11]), it is also known that the minimiser of such cost is $f_N = h^\omega_{V_N}$. The resulting cost, when plugged into (4.1) and divided by $N^{d-2}\log N$, yields

$$\liminf_{N\to\infty} \frac{1}{N^{d-2}\log N}\log \mathbb{P}^\omega[\mathcal{W}_N^+] \geq -d\overline{g}_{\mathbb{Q}} \limsup_{N\to\infty} \frac{\mathrm{Cap}^\omega(V_N)}{N^{d-2}} \overset{(2.11)}{=} -d\overline{g}_{\mathbb{Q}}\mathrm{Cap}^{a_{\mathbb{Q}}}(V)\,, \tag{4.5}$$

having used in the first step the definition of $\overline{g}_{\mathbb{Q}}$ in (3.2). However, this bound is not sharp, as shown by (3.1). This is due to the fact that the shift is not optimal, because it is considering the spins of the field over V_N as independent entries, neglecting the correlations between them, given by $g^\omega(\cdot,\cdot)$. Due to these correlations, it suffices to shift the field by a level which still has the same order $\sqrt{\log N}$ as the previous shift, given by (4.2), but with a lower constant, i.e. $\sqrt{4\overline{g}_{\mathbb{Q}}\log N}$. Hence, by shifting the Gaussian free field by the field $h^\omega_{V_N}\sqrt{4\overline{g}_{\mathbb{Q}}\log N}$, the sharp lower bound in (3.1) can be attained (see Section 4.1 of [6] for the details). In particular, it sufficies to shift the field on boxes of side lenght $L = o(N)$, with $L \to \infty$ as $N \to \infty$, by the first order behaviour of the expected maximum of the field over such mesoscopic boxes. This idea is also exploited in the proof of the upper bound, which we sketch in the next subsection.

4.2 Sketch of proof for the upper bound

The proof for the upper bound in (3.1) exploits the idea that it sufficies for spins on mesoscopic disjoint boxes with side length $L = o(N)$ to be above the first order behaviour of $\mathbb{E}^\omega[\max_{x\in V_L}\varphi_x]$ (see (4.2)) to lift the whole field over V_N above the zero level. It also uses the decomposition of the field in harmonic average and local field over these boxes, as in (2.13) and (2.14). Imposing that $L \to \infty$ as $N \to \infty$, since the harmonic average has low variance as L grows with N, then the maximal fluctuation of the field over the box is asymptotically entirely captured by the local field.

More precisely, in Section 4.2 of [6], disjoint boxes $U_{z_1} = B(z_1, KL), \dots, U_{z_{|\mathcal{C}_N|}} = B(z_{|\mathcal{C}_N|}, KL)$ contained in V_N, for $K \geq 100$ fixed, are considered, where $\mathcal{C}_N$ is the set of their centers, and the distance between z and z' is $|z-z'|= 4KL$ for $z, z' \in \mathcal{C}_N$ with $z \neq z'$, so that $|\mathcal{C}_N| \asymp (N/L)^d$. Given any of those boxes U, the decomposition $\phi = \xi^{\omega,U} + \psi^{\omega,U}$ in harmonic average and local field with respect to U is considered. (4.2) is then extended as a result on the first order behaviour of the maximum of ψ^{ω,U_z} over $B_z = B(z, L) \subseteq U_z$, uniformly over all $z \in \mathcal{C}_N$. In particular, it is shown that

$$\lim_{N\to\infty} \sup_{z\in\mathcal{C}_N} \frac{\mathbb{E}^\omega[\max_{x\in B_z}\psi^{\omega,U_z}_x]}{\sqrt{2d\overline{g}_{\mathbb{Q}}\log L}} = 1 \tag{4.6}$$

(note that this also gives the behaviour of the expected minimum of ψ^{ω,U_z} over B_z since the field is centered). For $\delta \in (0,1)$, the concept of *good* box is defined, as any box B_z,

$z \in \mathcal{C}_N$, where there exists a vertex where the local field is above the first-order behaviour of $\mathbb{E}^\omega[\min_{x\in B_z} \psi_x^{\omega,U_z}]$. That is, B_z is good if

$$\min_{x\in B_z} \psi_x^{\omega,U_z} \le -\sqrt{2d\overline{g}_{\mathbb{Q}}(1-\delta)\log L}\,. \tag{4.7}$$

By (4.6), most of the boxes are good. Note that, by (2.14), under the hard wall event, a good box necessarily contains a point where its harmonic average exceeds $\sqrt{2d\overline{g}_{\mathbb{Q}}(1-\delta)\log L}$, i.e.

$$\mathcal{W}_N^+ \subseteq \bigcap_{B_z \text{ good}} \{\max_{x\in B_z} \xi_x^{\omega,U_z} \ge \sqrt{2d\overline{g}_{\mathbb{Q}}(1-\delta)\log L}\}\,. \tag{4.8}$$

Since the boxes are disjoint, for two different boxes B_z and $B_{z'}$ and for $x \in B_z$ and $x' \in B_{z'}$, we have that ξ_x^{ω,U_z} and $\xi_{x'}^{\omega,U_{z'}}$ are independent random variables, by the spatial Markov property of the field (see (2.15)). We now consider a linear combination of the harmonic averages computed at the points where the maxima are attained in (4.8), namely

$$Z_{\widetilde{f}}^\omega = \sum_{z\in\mathcal{C}_N} \nu^\omega(z)\xi_{\widetilde{f}(z)}^{\omega,U_z} \tag{4.9}$$

where $\widetilde{f}(z) = \operatorname{argmax}_{x\in B_z}\xi^{\omega,U_z}$, $\nu^\omega(z) \ge 0$, $z \in \mathcal{C}_N$, and still get a Gaussian random variable $Z_{\widetilde{f}}^\omega$. Taking ν^ω in (4.9) to be a probability measure on $\mathcal{C}_N$ one has

$$\bigcap_{B \text{ good}} \{\max_{x\in B} \xi_x^{\omega,B} \ge \sqrt{2d\overline{g}_{\mathbb{Q}}(1-\delta)\log L}\} \subseteq \{Z_{\widetilde{f}}^\omega \ge \sqrt{2d\overline{g}_{\mathbb{Q}}\log L}\} \subseteq \{\sup_{f\in\mathcal{F}} Z_f^\omega \ge \sqrt{2d\overline{g}_{\mathbb{Q}}\log L}\}\,, \tag{4.10}$$

where $\mathcal{F} = \{f : \mathcal{C}_N \to \mathbb{Z}^d \,:\, f(z) \in B_z\,, z \in \mathcal{C}_N\}$. It is now possible to estimate the probability of the event in the right-hand side of the above display using Borell-TIS inequality. Due to the decay of the covariances of the field (see (2.4)), the side L of the boxes is set to be

$$L = \left\lceil \frac{N^{2/d}}{(\log N)^{1/2d}} \right\rceil\,. \tag{4.11}$$

This also yields

$$\lim_{N\to\infty} \frac{2d\overline{g}_{\mathbb{Q}}\log L}{4\overline{g}_{\mathbb{Q}}\log N} = 1\,, \tag{4.12}$$

cf. (3.3). Moreover, this choice of L makes sure that the combinatorial factor $2^{|\mathcal{C}_N|} \asymp 2^{((N/L)^d)}$ due to the position of the good boxes is of order $\exp\{o(N^{d-2}\log N)\}$, namely it is negligible when one is looking for an upper bound for (3.1). Hence, with this choice of L, (4.8) and (4.10) yield

$$\limsup_{N\to\infty} \frac{1}{N^{d-2}\log N} \log \mathbb{P}^\omega[\mathcal{W}_N^+] \le \limsup_{N\to\infty} \frac{1}{N^{d-2}\log N} \log \mathbb{P}^\omega[\sup_{f\in\mathcal{F}} Z_f^\omega \ge \sqrt{2d\overline{g}_{\mathbb{Q}}\log L}] \tag{4.13}$$

and by Borell-TIS inequality, applied after having studied $\mathbb{E}^\omega[\sup_{f\in\mathcal{F}} Z_f^\omega]$ and $\mathrm{Var}^\omega(Z_f^\omega)$ for $f \in \mathcal{F}$, we get

(4.14)
$$\limsup_{N\to\infty} \frac{1}{N^{d-2}\log N} \log \mathbb{P}^\omega[\sup_{f\in\mathcal{F}} Z_f^\omega \geq \sqrt{2d\overline{g}_{\mathbb{Q}} \log L}] \leq -2\overline{g}_{\mathbb{Q}}(1-\delta) \liminf_{N\to\infty} \frac{\mathrm{Cap}^\omega(\bigcup_{B \text{ good}} B)}{N^{d-2}} .$$

Using that most of the boxes are good, it is now possible to prove a solidification result, i.e.

$$\lim_{N\to\infty} \frac{\mathrm{Cap}^\omega(\bigcup_{B \text{ good}} B)}{\mathrm{Cap}^\omega(V_N)} = 1 . \tag{4.15}$$

Plugging this into (4.14) we get by (4.13)

(4.16)
$$\limsup_{N\to\infty} \frac{1}{N^{d-2}\log N} \log \mathbb{P}^\omega[\mathcal{W}_N^+] \leq -2\overline{g}_{\mathbb{Q}}(1-\delta) \liminf_{N\to\infty} \frac{\mathrm{Cap}^\omega(\bigcup_{B \text{ good}} B)}{N^{d-2}} \overset{(2.11)}{=} -2\overline{g}_{\mathbb{Q}}(1-\delta)\mathrm{Cap}^{a_{\mathbb{Q}}}(V)$$

and we get the upper bound for (3.1) by sending $\delta \to 0$.

References

[1] M.T. Barlow, "Random Walks and Heat Kernels on Graphs". Lecture note series. Cambridge University Press, 2017.

[2] M. Biskup, *Recent progress on the Random Conductance Model.* Probability Surveys, 8(none): 294–373, 2011.

[3] E. Bolthausen,J.-D. Deuschel, and O. Zeitouni, *Entropic Repulsion of the Lattice Free Field.* Communications in Mathematical Physics, 170(2): 417–443, 1995.

[4] A. Chiarini, A. Cipriani, and R.S. Hazra, *Extremes of the supercritical Gaussian Free Field.* Alea : Latin American Journal of Probability and Mathematical Statistics, 13(2): 711–724, 2016.

[5] A. Chiarini and M. Nitzschner, *Disconnection and entropic repulsion for the harmonic crystal with random conductances.* Communications in Mathematical Physics, 386: 1685–1745, 2020.

[6] Alberto Chiarini and Emanuele Pasqui, *Hardwall repulsion for the discrete gaussian free field in random environment on $\mathbb{Z}^d$, $d \geq 3$.* arXiv preprint arXiv:2510.24562, 2025.

[7] J.-D. Deuschel and G. Giacomin, *Entropic repulsion for the free field: Pathwise Characterization in $d \geq 3$.* Communications in Mathematical Physics, 206(2): 447–462, 1999.

[8] S. Neukamm, M. Schäffner, and A. Schlömerkemper, *Stochastich omogenization of nonconvex discrete energies with degenerate growth.* SIAM Journal on Mathematical Analysis, 49(3): 1761–1809, 2017.

[9] S.C. Port and C.J. Stone, "Brownian motion and classical potential theory". Probability and Mathematical Statistics. Academic Press, New York, 1978.

[10] A.-S. Sznitman, "Topics in Occupation Times and Gaussian Free Fields". Zurich lectures in advanced mathematics. European Mathematical Society, 2012.

[11] A.-S. Sznitman, *Disconnection, random walks, and random interlacements.* Probab. Theory Relat. Fields,167(1-2): 1–44, 2017.

Springer Theory, Hecke Categories, and Motivic Centers: an expository introduction through the example of SL_2

Runlei Xiao (*)

Abstract. These notes explain how Weyl-group symmetries arise from geometry and how they enter a categorical-center description of equivariant motivic sheaves on a reductive group. The running example is $G = SL_2$. We begin with the flag variety, the Bruhat decomposition, and the Grothendieck–Springer resolution. Over the regular semisimple locus, the resolution is a finite étale cover with deck-transformation group the Weyl group. The Steinberg variety extends this symmetry across the singular locus and leads to the Springer sheaf and its motivic analogue. We then introduce convolution, the universal Hecke category, the Harish–Chandra transform, and the categorical center. The final sections outline the motivic center theorem and the comonadic mechanism behind its proof.
The purpose is expository. Some cohomological shifts, Tate twists, finiteness hypotheses, and choices of motivic formalism are suppressed in the first half of the notes and restored only when they are essential to the main statement.

1 Why study representation theory through geometry?

In classical representation theory, a group Γ acts on a vector space V. A representation is therefore a homomorphism

$$\rho\colon \Gamma \longrightarrow GL(V).$$

Geometric representation theory replaces the single vector space V by a family of linear objects parametrized by a space. Depending on the context, these objects may be constructible sheaves, complexes of sheaves, D-modules, or motives. A group action is then encoded by geometric correspondences between parameter spaces.

The slogan for these notes is:

> *Weyl-group representations arise from the geometry of flags compatible with a matrix, and categorical centers organize the resulting symmetries.*

(*)Ph.D. course, Università di Padova, Dip. Matematica, via Trieste 63, I-35121 Padova, Italy. E-mail: runlei.xiao@phd.unipd.it . Seminar held on 12 February 2026.

There are three levels in the story.

Level 1. Geometry: resolutions of singular spaces, flag varieties, and the Steinberg variety.

Level 2. Linearization: sheaves, Chow groups, or motives turn geometric correspondences into linear or categorical operators.

Level 3. Categorical structure: convolution and centers identify which operators commute with the Hecke action.

We first work concretely with SL_2. This example already contains the identity element and the nontrivial reflection of the Weyl group, the two Bruhat cells, the two sheets of the Grothendieck–Springer cover, and the two basic components of the Steinberg variety.

2 The running example: SL_2, flags, and the Weyl group

Let k be a field. To avoid inseparability issues in the elementary discussion, one may first assume that k is algebraically closed and that $\mathrm{char}(k) \neq 2$. Set

$$G = SL_2(k) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \;\middle|\; ad - bc = 1 \right\}.$$

Let $B \subset G$ be the subgroup of upper-triangular matrices and let $T \subset B$ be the subgroup of diagonal matrices:

$$B = \left\{ \begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix} \right\}, \qquad T = \left\{ \begin{pmatrix} t & 0 \\ 0 & t^{-1} \end{pmatrix} \right\}.$$

The subgroup B is a Borel subgroup and T is a maximal torus.

2.1 The flag variety

A complete flag in k^2 is simply a line $L \subset k^2$. The group G acts transitively on the set of lines, and the stabilizer of the standard line ke_1 is B. Hence

$$G/B \simeq \mathbf{P}^1.$$

This is the first appearance of the flag variety. A point gB corresponds to the line $g(ke_1)$.

For a general split reductive group G, the quotient G/B parametrizes Borel subgroups of G: the point gB corresponds to gBg^{-1}. Thus one can regard G/B either as a space of flags or as a moduli space of Borel subgroups.

2.2 The Weyl group

The Weyl group of (G, T) is

$$W = N_G(T)/T.$$

For SL_2, the normalizer contains the matrix

$$s = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

which interchanges the two diagonal entries of an element of T. Therefore

$$W \simeq \mathbb{Z}/2\mathbb{Z} = \{e, s\}.$$

The nontrivial element acts on T by

$$\begin{pmatrix} t & 0 \\ 0 & t^{-1} \end{pmatrix} \longmapsto \begin{pmatrix} t^{-1} & 0 \\ 0 & t \end{pmatrix}.$$

For GL_n, the same construction gives $W \simeq S_n$, acting by permuting the diagonal entries.

2.3 Bruhat decomposition for SL_2

Proposition 2.1 (Bruhat decomposition for SL_2) *There is a disjoint union*

$$SL_2 = B \sqcup BsB.$$

Equivalently,

$$G = \bigsqcup_{w \in W} BwB.$$

Proof. Let

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2.$$

If $c = 0$, then $g \in B$. Suppose $c \neq 0$. A direct factorization gives

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & a/c \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} c & d \\ 0 & c^{-1} \end{pmatrix},$$

after a harmless adjustment of signs depending on the chosen representative of s. Thus $g \in BsB$. □

The geometric version is the orbit decomposition of $G/B \times G/B$ under the diagonal G-action. For SL_2, a pair of lines (L_1, L_2) is either equal or distinct. Hence there are two orbits:

$$\Delta = \{(L, L)\} \qquad \text{and} \qquad (\mathbf{P}^1 \times \mathbf{P}^1) \setminus \Delta.$$

They are indexed by e and s in the Weyl group. This simple observation is the geometric seed of the Hecke category.

3 The Grothendieck–Springer resolution

3.1 Definition and moduli interpretation

For a connected reductive group G, define

$$\widetilde{G} = \{(g, B') \mid g \in G,\ B' \subset G \text{ a Borel subgroup, and } g \in B'\}.$$

The projection

$$\pi\colon \widetilde{G} \longrightarrow G, \qquad (g, B') \longmapsto g,$$

is the group-theoretic Grothendieck–Springer resolution. Using a fixed Borel B, one may also write

$$\widetilde{G} \simeq G \times^B B.$$

A point of the fibre $\pi^{-1}(g)$ is a Borel subgroup containing g.

For $G = SL_2$, a Borel subgroup is the stabilizer of a line in k^2. Consequently,

$$\widetilde{G} \simeq \{(g, L) \in SL_2 \times \mathbf{P}^1 \mid gL = L\}.$$

Thus the resolution remembers an invariant line of the matrix g.

3.2 The regular semisimple locus

An element $g \in SL_2$ has characteristic polynomial

$$\chi_g(x) = x^2 - \operatorname{tr}(g)x + 1.$$

It is regular semisimple when this polynomial has two distinct roots. Over an algebraically closed field of characteristic different from 2, this is equivalent to

$$\operatorname{tr}(g)^2 - 4 \neq 0.$$

Write $G^{\mathrm{rs}} \subset G$ for the regular semisimple locus.

A regular semisimple matrix in SL_2 has two distinct eigenlines. Each eigenline determines a Borel subgroup containing the matrix. Therefore the fibre of π over $g \in G^{\mathrm{rs}}$ has two points. The nontrivial Weyl-group element exchanges them.

Theorem 3.1 *The restriction*

$$\pi^{\mathrm{rs}}\colon \widetilde{G}^{\mathrm{rs}} \longrightarrow G^{\mathrm{rs}}$$

is a finite étale W-cover. For $G = SL_2$, it is a double cover, and its deck transformation exchanges the two eigenlines.

This theorem gives a geometric explanation of the Weyl group: it is the ambiguity in choosing and ordering the eigen-data of a regular semisimple element.

Remark 3.2 For GL_n, a regular semisimple matrix has n distinct eigenlines. A complete invariant flag is equivalent to an ordering of these eigenlines, so the fibre has $n!$ points and the symmetric group S_n acts simply transitively.

4 From the cover to Springer theory

The Weyl-group action is clear over G^{rs}, but many interesting elements of G are not regular semisimple. The central problem is to extend the symmetry across the singular locus.

4.1 The unipotent and nilpotent resolutions

Let $\mathcal{U} \subset G$ be the unipotent locus. Restricting the Grothendieck–Springer resolution gives

$$p{:}\widetilde{\mathcal{U}} := \widetilde{G} \times_G \mathcal{U} \longrightarrow \mathcal{U}.$$

In the Lie algebra $\mathfrak{g}$, one similarly has the nilpotent cone $\mathcal{N}$ and the Springer resolution

$$p_{\mathfrak{g}}{:}\widetilde{\mathcal{N}} \longrightarrow \mathcal{N}.$$

Under suitable hypotheses, a Springer isomorphism relates the unipotent and nilpotent pictures. The Lie-algebra version is often easier to visualize, while the group version is better adapted to character sheaves and conjugation-equivariant categories.

For SL_2, the fibre over the identity element in the group picture, or over 0 in the Lie algebra picture, is the full flag variety

$$G/B \simeq \mathbf{P}^1.$$

Over a nontrivial regular unipotent element, there is a unique compatible line, so the fibre is a point.

4.2 The Springer sheaf and the Springer motive

In a sheaf-theoretic setting, the Springer object is obtained by pushing forward the constant sheaf which is exceptional pullback from the structure map $\widetilde{N} \to \operatorname{Spec} k$ along the resolution:

$$\Sigma_{\mathrm{sh}} := p_* f^! \mathbb{Q}_S,$$

Springer theory constructs an action

$$\mathbb{Q}[W] \longrightarrow \mathrm{End}(\mathrm{Spr}_{\mathrm{sh}}).$$

The motivic analogue replaces the constant sheaf by the motivic unit:

$$\Sigma_{\mathrm{mot}} := p_* f^! 1_S \qquad \text{in} \qquad \mathrm{DM}(\mathcal{U}/G)$$

or, after extension to G/G, by the corresponding Springer motive Σ. A motive may be viewed informally as a universal geometric object whose realization in Betti, ℓ-adic, or other cohomology theories recovers familiar sheaf-theoretic objects. Constructing the Weyl-group action motivically therefore produces compatible actions after realization.

5 The Steinberg variety and convolution

5.1 Definition

The Steinberg variety is the fibre product

$$\mathrm{St}_{\mathcal{U}} := \widetilde{\mathcal{U}} \times_{\mathcal{U}} \widetilde{\mathcal{U}}.$$

A point consists of a unipotent element together with two Borel subgroups containing it. In the Lie-algebra setting, one writes

$$\mathrm{St}_{\mathcal{N}} := \widetilde{\mathcal{N}} \times_{\mathcal{N}} \widetilde{\mathcal{N}}.$$

Over the regular semisimple locus, the analogous fibre product decomposes as the union of the graphs of the Weyl-group action:

$$\widetilde{G}^{\mathrm{rs}} \times_{G^{\mathrm{rs}}} \widetilde{G}^{\mathrm{rs}} \simeq \bigsqcup_{w \in W} \Gamma_w.$$

The closures of these graphs inside the global Steinberg space produce cycles indexed by W.

For SL_2, there are two basic pieces: the diagonal component corresponding to e, and the component corresponding to the reflection s.

5.2 Convolution

Convolution is geometric composition. Consider the triple fibre product

$$\widetilde{\mathcal{U}} \times_{\mathcal{U}} \widetilde{\mathcal{U}} \times_{\mathcal{U}} \widetilde{\mathcal{U}}$$

with projections p_{12}, p_{23}, p_{13} to the indicated pairs. For suitable classes a, b on the Steinberg variety, define

$$a * b = (p_{13})_*(p_{12}^* a \cdot p_{23}^* b).$$

The middle flag is composed away. This is the geometric analogue of multiplying kernels or composing relations.

Under standard hypotheses, the top-dimensional Chow group becomes a convolution algebra, and one has an isomorphism

$$\mathrm{CH}_{\mathrm{top}}(\mathrm{St}_{\mathcal{N}})_{\mathbb{Q}} \simeq \mathbb{Q}[W].$$

This description turns geometric cycles into Weyl-group operators on the Springer object.

5.3 Two geometric strategies for the motivic action

The motivic construction in the present project is approached in two related ways.

(1) **Fourier-transform method.** Reduce the group case to the Lie-algebra case by a Springer isomorphism and apply homogeneous Fourier transformation, following the classical strategy for Weyl-group actions on the Springer sheaf. One then checks that the relevant cycles and transformations are equivariant in the required motivic setting.

(2) **Specialization method.** Start with the graphs Γ_w over the regular semisimple locus and specialize their classes to the Steinberg variety. The goal is to prove that this specialization is compatible with convolution. For SL_2, the geometry reduces to the two elements $e, s \in W$ and provides a useful test case.

6 From geometry to categories

6.1 Correspondences as generalized maps

A correspondence from X to Y is a diagram

$$X \xleftarrow{p} Z \xrightarrow{q} Y.$$

A six-functor formalism assigns to such a diagram a functor between categories of sheaves or motives. Schematically, one obtains a lax symmetric monoidal functor

$$\mathrm{DM}_!^*\colon \mathrm{Corr}(\mathcal{M}) \longrightarrow \mathrm{Pr}^{\mathrm{L}},$$

where $\mathcal{M}$ is a suitable category of schemes or stacks. Depending on the variance, the correspondence may act by a composite such as $q_! p^*$ or $q_* p^!$.

The point is that a geometric relation produces an operator. Fibre products compose relations, and proper pushforward turns this composition into convolution.

6.2 Quotient stacks

If an algebraic group G acts on a space X, the quotient stack X/G remembers both the points of X and their stabilizers. The category

$$\mathrm{DM}(X/G)$$

should be thought of as the category of G-equivariant motives on X. In particular,

$$\mathrm{DM}(G/G)$$

is the category of motives on G equivariant for conjugation.

The multiplication map $G \times G \to G$ gives $\mathrm{DM}(G/G)$ a convolution product. Thus the geometry of conjugacy naturally produces a monoidal category.

7 The universal Hecke category

The flag variety produces a second convolution category. In the form relevant to the center theorem, define

$$\mathcal{H}_G := \mathrm{DM}(BB \times_{BG \times BT} BB) \simeq \mathrm{DM}((G/B \times_{BT} G/B)/G).$$

Here BG, BB, and BT are classifying stacks. The fibre product records two Borel reductions together with their torus data. The category $\mathcal{H}_G$ is called the finite universal Hecke category.

For a first reading, it is useful to suppress the BT-refinement and remember only the simpler space

$$(G/B \times_{BT} G/B)/G.$$

Its strata are indexed by the Weyl group. For SL_2, the diagonal and its complement give the two basic Hecke objects corresponding to e and s.

Convolution is induced by the three-fold flag space:

$$\begin{array}{ccccc} & & (G/B)^3/G & & \\ & \swarrow^{p_{12}} & \downarrow{\scriptstyle p_{13}} & \searrow^{p_{23}} & \\ (G/B)^2/G & & (G/B)^2/G & & (G/B)^2/G. \end{array}$$

Given $K, L \in \mathcal{H}_G$, one has schematically

$$K * L = (p_{13})_!(p_{12}^* K \otimes p_{23}^* L).$$

The universal Hecke category is naturally a module over a larger monoidal category $\mathcal{A}_G$ encoding endomorphisms of the underlying correspondence. This extra action is needed for the correct relative notion of center.

8 What is a categorical center?

For an ordinary algebra A, the center is

$$Z(A) = \{a \in A \mid ab = ba \text{ for all } b \in A\}.$$

An element of $Z(A)$ acts on every A-module in a way compatible with the A-action.

A categorical center replaces elements by endofunctors and equality by coherent isomorphisms. If a monoidal category $\mathcal{A}$ acts on a category $\mathcal{M}$, the relative center is represented by the category of $\mathcal{A}$-linear endomorphisms of $\mathcal{M}$:

$$\mathfrak{Z}_{\mathcal{A}}(\mathcal{M}) := \mathrm{End}_{\mathcal{A}}(\mathcal{M}).$$

In our setting,

$$\mathfrak{Z}_G(\mathcal{H}_G) := \mathrm{End}_{\mathcal{A}_G}(\mathcal{H}_G).$$

The main conceptual question is therefore:

> Which geometric objects on G/G produce endofunctors of the universal Hecke category that commute with the full Hecke action?

The answer is given by the Harish–Chandra transform.

9 The Harish–Chandra transform

There is a geometric correspondence connecting conjugacy classes in G with flags:

$$\mathrm{DM}(G/G) \underset{\chi}{\overset{\mathfrak{hc}}{\rightleftarrows}} \mathcal{H}_G.$$

The functor $\mathfrak{hc}$ is the Harish–Chandra transform and χ is its right adjoint, often called the character functor. At the geometric level, the correspondence records an element of G together with a Borel subgroup compatible with it. Thus it is a categorical version of the same incidence relation that defines the Grothendieck–Springer resolution.

The exact pull–push formula depends on the chosen six-functor formalism. Schematically, if

$$G/G \xleftarrow{q} \mathcal{X} \xrightarrow{r} BB \times_{BG\times BT} BB$$

is the Harish–Chandra correspondence, then

$$\mathfrak{hc} \simeq r_! q^*, \qquad \chi \simeq q_* r^!.$$

The adjunction gives a unit transformation

$$\eta\colon \mathrm{id}_{\mathrm{DM}(G/G)} \longrightarrow \chi \circ \mathfrak{hc}.$$

A central geometric fact is that the composite $\chi \circ \mathfrak{hc}$ is controlled by the Springer motive.

10 The motivic center theorem

10.1 Statement

Theorem 10.1 (Motivic categorical center) *Let G be a split reductive group over a base scheme S, and work in a motivic six-functor formalism satisfying the required descent, finiteness, and duality assumptions. Then the Harish–Chandra transform induces a monoidal equivalence*

$$\mathrm{DM}(G/G) \simeq \mathrm{End}_{\mathcal{A}_G}(\mathcal{H}_G).$$

Informally, the theorem says:

> *Conjugation-equivariant motives on G are the central operations on the universal Hecke category.*

The left-hand side is built from conjugacy geometry; the right-hand side is built from the flag variety. Springer theory provides the bridge.

10.2 The Springer kernel

Let $e\colon S \to G$ be the identity section and let

$$\delta_e := e_* \mathbf{1}_S \in \mathrm{DM}(G/G)$$

be the convolution unit. Let Σ denote the Springer motive on G/G. The unit of the Harish–Chandra adjunction corresponds to a natural map

$$\alpha\colon \delta_e \longrightarrow \Sigma.$$

Moreover, convolution with Σ describes the adjunction monad or comonad:

$$\chi \circ \mathfrak{hc} \simeq \Sigma * (-).$$

Thus the abstract adjunction is represented by a concrete geometric object.

The project constructs a map

$$\beta\colon \Sigma \longrightarrow \delta_e$$

whose composite with α is governed by the motivic Euler characteristic of the flag variety. In motivic stable homotopy theory, one obtains schematically

$$e^*(\beta \circ \alpha) = (-1)^{\dim(G/B)} \chi^{\mathrm{cat}}(G/B).$$

For a split reductive group, the relevant calculation has the form

$$\chi^{\mathrm{cat}}(G/B) = \frac{|W|}{2}(\langle -1 \rangle + \langle 1 \rangle).$$

After passing to rational motives, or equivalently after the required hyperbolic localization, this scalar is invertible. Therefore α admits a section.

For SL_2, one has $G/B \simeq \mathbf{P}^1$ and $|W| = 2$. The calculation is the motivic refinement of the elementary fact

$$\chi(\mathbf{P}^1) = 2.$$

10.3 Comonadic reconstruction

Set

$$\mathbb{S} := \mathfrak{hc} \circ \chi.$$

This is a comonad on $\mathcal{H}_G$. The splitting of the adjunction unit implies that objects of $\mathrm{DM}(G/G)$ can be reconstructed from their Harish–Chandra transforms together with the $\mathbb{S}$-coaction.

The technical input is the following variant of Barr–Beck–Lurie.

Theorem 10.2 (Unipotent-retraction comonadicity) *Let*

$$F\colon \mathcal{C} \rightleftarrows \mathcal{D} : G$$

be an adjunction, with F left adjoint to G. Assume that $\mathcal{C}$ and $\mathcal{D}$ admit the required totalizations. Suppose the unit

$$\eta\colon \mathrm{id}_{\mathcal{C}} \longrightarrow GF$$

admits a retraction $s\colon GF \to \mathrm{id}_{\mathcal{C}}$ such that

$$(s \circ \eta)^n \simeq \mathrm{id}_{\mathcal{C}}$$

for some $n \geq 1$. Then F is comonadic and induces an equivalence

$$\mathcal{C} \simeq \mathrm{LComod}_{FG}(\mathcal{D}).$$

Applying this to the Harish–Chandra adjunction yields

$$\mathrm{DM}(G/G) \simeq \mathrm{LComod}_{\mathbb{S}}(\mathcal{H}_G).$$

10.4 The same comonad on the center

The final step is to identify the categorical center with the same comodule category:

$$\mathrm{End}_{\mathcal{A}_G}(\mathcal{H}_G) \simeq \mathrm{LComod}_{\mathbb{S}}(\mathcal{H}_G).$$

The comonad $\mathbb{S}$ is represented by a coalgebra object $S_2 \in \mathcal{A}_G$. An $\mathcal{A}_G$-linear endofunctor of $\mathcal{H}_G$ is determined by its value on the unit object $\mathbf{1}_{\mathcal{H}_G}$, and this value carries a canonical S_2-coaction. Combining the two comonadic descriptions gives

$$\mathrm{DM}(G/G) \simeq \mathrm{LComod}_{\mathbb{S}}(\mathcal{H}_G) \simeq \mathrm{End}_{\mathcal{A}_G}(\mathcal{H}_G).$$

11 A formal categorical mechanism

The proof above is an instance of a more general pattern. Let

$$F\colon \mathcal{C}^{\otimes} \longrightarrow \mathcal{D}^{\otimes}$$

be a lax symmetric monoidal functor between symmetric monoidal $(\infty, 2)$-categories. Let A and E be self-dual E_1-algebra objects in $\mathcal{C}$, and let $f\colon A \to E$ admit a right adjoint. The action of E on itself gives a map

$$E \longrightarrow \mathrm{End}(E).$$

Suppose that the composite

$$A \longrightarrow E \longrightarrow \mathrm{End}(E)$$

factors through the center of the $\mathrm{End}(E)$-module E. After applying F, one obtains a map

$$F(A) \longrightarrow \mathrm{End}_{F(\mathrm{End}(E))}(F(E)).$$

Under the geometric and comonadic hypotheses appearing above, this map is an equivalence.

This suggests a relative center attached to any lax monoidal functor.

Definition 11.1 Let $F\colon \mathcal{C}^{\otimes} \to \mathcal{D}^{\otimes}$ be lax monoidal and let $M \in \mathcal{C}$. Define the center of M relative to F by

$$\mathfrak{Z}_F(M) := \mathrm{End}_{F(\mathrm{End}_{\mathcal{C}}(M))}(F(M)).$$

In the motivic application, $\mathcal{C}$ is a correspondence category, F is the motivic six-functor formalism, A is represented by the group object G/G, and E is represented by the flag correspondence underlying the universal Hecke category.

12 What is proved geometrically, and what is formal?

It is useful to separate the proof into two parts.

Geometric input

- the Grothendieck–Springer resolution and its W-cover over the regular semisimple locus;
- the Steinberg variety and its convolution cycles;
- the identification of $\chi \circ \mathfrak{hc}$ with convolution by the Springer motive;
- the construction of the maps $\delta_e \to \Sigma \to \delta_e$;
- the motivic Euler-characteristic calculation for G/B;
- the motivic Weyl-group action on Σ.

Formal categorical input

- six-functor formalisms send correspondences to functors;
- convolution becomes composition of kernels;
- the split or unipotently split unit implies comonadicity;
- the relative center is reconstructed from the value of an endofunctor on the unit;
- both sides of the center theorem are categories of comodules over the same comonad.

This separation is important: the categorical machine does not replace the geometry. It organizes the geometric information so that the same proof can be transported to other sheaf theories and motivic settings.

13 Outlook: twisted and monodromic motives

The center theorem motivates a broader study of equivariance with nontrivial monodromy. Let an algebraic group G act on a scheme X, and let $L \in \mathrm{DM}(G)$ be an invertible character sheaf, meaning schematically

$$m^*L \simeq L \boxtimes L, \qquad e^*L \simeq \mathbf{1}.$$

If ω_G denotes the dualizing object of G, one expects a category of L-twisted equivariant motives of the form

$$\mathrm{DM}(X/_L G) := \mathrm{LMod}_{L\otimes\omega_G}(\mathrm{DM}(X)).$$

For a torus, character sheaves can be produced from finite tame covers and their idempotent summands. Passing to a universal tame cover suggests a motivic category of monodromic sheaves.

Two natural goals are:

(1) construct a six-functor formalism for monodromic motives;

(2) prove hyperbolic localization for the monodromic subcategory.

These questions extend the same guiding principle of the present notes: geometric correspondences should reveal hidden algebraic and categorical symmetries.

References

[1] P.N. Achar, A. Henderson, D. Juteau, and S. Riche, *Weyl group actions on the Springer sheaf.* Proc. London Math. Soc. 108 (2014), no. 6, 1501–1528.

[2] R. Bezrukavnikov, A. Ionov, K. Tolmachov, and Y. Varshavsky, *Equivariant derived category of a reductive group as a categorical center.* arXiv:2305.02980.

[3] N. Chriss and V. Ginzburg, "Representation Theory and Complex Geometry". Birkhäuser, 1997.

[4] F. Déglise, F. Jin, and A.A. Khan, *Fundamental classes in motivic homotopy theory.* J. Eur. Math. Soc. 23 (2021), no. 12, 3935–3993.

[5] M. Hoyois, *The six operations in equivariant motivic homotopy theory.* Adv. Math. 305 (2017), 197–279.

[6] J. Lurie, "Higher Algebra". 2017 manuscript.

Profinite groups and probability

Nowras Naufel (*)

Abstract. The main aim of these notes is to give a very gentle introduction to profinite groups, while also attempting to illustrate the breadth of this class of topological groups by means of examples coming from quite diferente branches of mathematics. Towards the end, we take a peek into one of the many ways in which probability can play a role in raising interesting questions about profinite groups and in providing means to solve them.

1 Introduction

Profinite groups have, at this point, a long tradition in group theory. They form a class of topological groups which naturally appears in many distinct areas of mathematics.

We name a few. In Galois theory, work of Krull in the early 20th century showed that Galois groups of infinite Galois extensions are profinite and a lot of general theory was developed in order to deepen our understanding of these infinite extensions (many of which are notable, such as the algebraic closure of the rational numbers). In algebraic geometry, the (classical) theory of Brauer groups, which arises as an object to classify central simple algebras over a field, is associated to profinite groups, since the Brauer group of a field is isomorphic to the second cohomology group of certain profinite groups; this relation is essential in proving various properties of the Brauer group. In the study of Lie groups over p-adic fields, where the interplay between profinite groups and non-archimedean analysis proved very fruitful in understanding this class of groups. As a final example, we mention their importance also in the theory of abstract groups; the program of understanding when a finitely generated residually finite group is determined by its finite quotients makes heavy use of the *profinite completion* of an abstract group, which is a profinite group encoding all of these finite quotients.

To summarise, the beauty of this theory is that, in its connection with many other fields, one can approach it through many differents prisms, merging together tools of algebra, topology, geometry and analysis.

In this short note, after introducing the basics of profinite groups and explaining some basic examples, we will highlight in particular the probabilistic tools that can come into play.

(*)Ph.D. course, Università di Padova, Dip. Matematica, via Trieste 63, I-35121 Padova, Italy. E-mail: naufel@math.unipd.it . Seminar held on 26 February 2026.

2 Profinite groups

In this section, we define our main object of study. First, recall that a group G with a topology τ is a *topological group* if the functions $(x, y) \mapsto xy$ (multiplication of the group) and $x \mapsto x^{-1}$ are continuous with respect to τ.

Examples abound: (i) any group with the discrete topology; (ii) $\mathrm{SL}_2(\mathbb{C})$ with the induced topology from $\mathbb{C}^4$; (iii) the circle S^1 with the induced topology from $\mathbb{C}$. All of these are Hausdorff; every group in (i) is *totally disconnected* (that is, its connected subsets are singletons) and the groups in (ii) and (iii) are connected; among those, only discrete finite groups and the circle are compact.

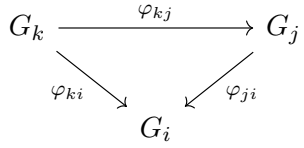
Definition 2.1 A topological group G is *profinite* if it is compact, Hausdorff and totally disconnected.

This is the first definition, and while it is very easy to state, it fails to convey *how* to study these groups. As such, we move on to the more involved and more useful definition.

Let $(I, \leq)$ be a *directed partially ordered* set, that is, for each $i, j \in I$, there exists $k \in I$ such that $i, j \leq k$. Suppose that there are finite groups G_i for each $i \in I$ and group homomorphisms $\varphi_{ji}\colon G_j \to G_i$ whenever $i \leq j$. Suppose further that the following two conditions are satisfied:

(i) $\varphi_{ii} = \mathrm{id}_{G_i}$ for each i;

(ii) whenever $i \leq j \leq k$, the diagram



commutes, i.e., the equality $\varphi_{ki} = \varphi_{ji}\varphi_{kj}$ is satisfied.

The triple $\mathcal{I} = (G_i, \varphi_{ji}, I)$ is an *inverse system of finite groups*. The *inverse limit* of $\mathcal{I}$ is the group defined as

$$\varprojlim_{i \in I} G_i := \{(g_i) \in \prod_{i \in I} G_i : \varphi_{ji}(g_j) = g_i \text{ whenever } i \leq j\} \subseteq \prod_{i \in I} G_i = \overline{G}.$$

It is the subset of 'nested sequences', since the entries are related by means of the family of homomorphisms in the inverse system. As stated in Definition 2.1, profinite groups are topological groups, so we will 'topologise' both of the groups above in an appropriate way.

First, we endow each G_i with the discrete topology, with which they become compact, Hausdorff and totally disconnected topological groups. Then, we endow $\overline{G}$ with the product topology, so $\overline{G}$ is also compact (by Tychonoff's theorem), Hausdorff and totally disconnected (and hence profinite by Definition 2.1).

Second, our aim is to prove that the inverse limit is a *closed* subgroup of $\overline{G}$. For each $i \in I$, let $\pi_i\colon \overline{G} \to G_i$ be the projection onto G_i and, for each $j \geq i$, let $\overline{\varphi}_{ji} = \varphi_{ji}\pi_j$; both are continuous maps. Again, for $j \geq i$, denote by L_{ji} the set

$$\{(g_i) \in \overline{G} : \pi_i((g_i)) = \overline{\varphi}_{ji}((g_i))\}$$

and define for each $i \in I$

$$L_i = \bigcap_{j \in I,\, i \leq j} L_{ji}\,.$$

Since $\overline{G}$ is a Hausdorff topological space, the sets L_{ji} are closed and thus so is L_i. Notice that

$$\varprojlim_{i \in I} G_i = \bigcap_{i \in I} L_i\,,$$

so the inverse limit is closed. Therefore, it inherits the properties of being compact, Hausdorff and totally disconnected. The reason we introduced the inverse limit construction is the following.

Theorem 2.2 ([RZ10, Theorem 1.1.12]) *Let G be a topological group. Then G is profinite if, and only if, it is an inverse limit of finite groups.*

The advantages of working with the inverse limit will soon become clear, as it naturally suggests certain definitions that come from the theory of finite groups.

Remarks 2.3

(i) The inverse limit construction above is a very particular case of a much more general construction. In general, one may replace 'finite groups' by sets, graphs, topological spaces, rings, algebras, among others, and appropriately replace 'group homomorphisms' by the correct morphisms according to the category in which one is working. All of these show up in the theory of profinite groups;

(ii) The attentive reader might wonder whether the inverse limit of finite groups is even a non-empty set. Indeed, in other categories, that is not necessarily the case; in [Wat72], Waterhouse gives an example of an inverse limit of non-empty sets which is empty. However, the inverse limit of compact Hausdorff non-empty spaces is *always* non-empty. Although this is a basic fact, it plays an important role in many proofs;

(iii) Profinite groups do *not* have a unique decomposition as an inverse limit.

Before carefully working on a few examples of profinite groups, we highlight a few important properties.

Proposition 2.4 *Let G be a profinite group. Then the identity element of G admits a fundamental system of neighborhoods $\mathcal{U}$ consisting of open normal subgroups U such that $\cap_{U \in \mathcal{U}} U = 1$, each G/U is a finite group and*

$$G = \varprojlim_{U \in \mathcal{U}} G/U\,.$$

This proposition roughly says that the inverse limit structure of a profinite group G is 'intrinsic', meaning that the finite groups that appear in the inverse system come from finite quotients of G.

The claim that each G/U is finite follows directly from the fact that U is an open normal subgroup in a compact topological group; indeed, notice that

$$G \smallsetminus U = \bigcup_{g \in T} gU \,,$$

where T is a set of representatives of the left cosets of U in G. Since U is open and 'multiplication by g' is a homeomorphism, then each gU is open and compactness tells us that $G \smallsetminus U$ is a union of finitely many left cosets, so U has finite index in G. Notice also that U is closed, so the existence of a fundamental system of neighborhoods consisting of clopen sets reflects the fact that G is totally disconnected.

We use the notation $\leq_c$ to indicate that a subgroup is *closed*. If $H \leq_c G$ and $N \trianglelefteq_c G$, then the groups H and G/N are also profinite and, from the inverse limit decomposition of G as in Proposition 2.4, we can write them as

$$H = \varprojlim_{U \in \mathcal{U}} HU/U$$

and

$$G/N = \varprojlim_{U \in \mathcal{U}} G/NU \,.$$

Finally, again as a consequence of Tychonoff's theorem, an arbitrary product of profinite groups is again profinite. These last three facts indicate that we are working with a reasonable class of groups, meaning that we remain within it after passing to subgroups or to quotients (while respecting the topology) or after taking products.

Recall that, for finite groups G, we define their order to be the cardinality $|G|$. The cardinality of a finite group often yields non-trivial information about it; important manifestations of this are the Sylow theorems. In general, for an infinite group G, there is no good definition of order, let alone a generalisation of Sylow theory. We will see now how the inverse limit allows us to reasonably define the order of a profinite group. We define *supernatural numbers* to be a formal product of the form

$$n = \prod_{p \text{ prime}} p^{m(p)} \,,$$

where $m(p) \in \mathbb{N} \cup \{\infty\}$. The notions of divisibility, of maximal common divisor and of lowest common multiple mimick the ones for natural numbers. If G is a profinite group and

$$G = \varprojlim_{U \in \mathcal{U}} G/U \,,$$

the *order* of G is

$$\#G = \operatorname{lcm}\{|G/U| \colon U \in \mathcal{U}\}$$

and, if $H \leq_c G$, the *index* of H in G is

$$[G : H] = \operatorname{lcm}\{|G/U : HU/U| \colon U \in \mathcal{U}\} \,.$$

We remark that these definitions do not depend on the inverse limit decomposition of G. A profinite group G is a *pro-p* group, for a prime p, if $\#G = p^{n(p)}$ or, equivalently, if G can be written as an inverse limit of finite p-groups.

Theorem 2.5 *Let G be a profinite group. If a prime p divides $\#G$, then G contains a Sylow p-subgroup.*

Here, a Sylow p-subgroup is understood as a closed subgroup P of G such that P is pro-p and p does not divide the index $[G : P]$. An element x of G is a *p-element* if $\#\overline{\langle x \rangle} = p^{n(p)}$. The reason we mention this theorem is that it is proven using the corresponding statement for finite groups and an inverse limit argument. A common occurrence is precisely that: 'lifting' a result from finite groups to profinite groups, by means of the fact that they are 'built' out of finite groups.

One final property of a profinite group G that we would like to highlight is, since it is compact and Hausdorff, then we can endow it with (normalised) Haar measure $\mu = \mu_G$ by a theorem of Haar and Weil. The adjective 'normalised' means that $\mu(G) = 1$, so it is in particular a probability measure. This fact will be essential for the next section.

The simplest examples of profinite groups are finite groups with the discrete topology. One non-example is the group of integers; profinite groups are either finite or uncountable (as a consequence of the Baire category theorem), so there is no topology on $\mathbb{Z}$ which turns it into a profinite group.

2.1 The pro-$\mathcal{C}$ completion

In this section, we use $\leq_f$ to denote a subgroup of finite index. For simplicity, let $\mathcal{C}$ be a class of finite groups closed under taking subgroups, quotients and finite direct products, and let G be any abstract group. Define

$$\mathcal{N}_{\mathcal{C}}(G) = \{N \trianglelefteq_f G : G/N \in \mathcal{C}\}.$$

This set is directed and partially ordered with respect to reverse inclusion. We define an inverse system as follows. To each $U \in \mathcal{N}_{\mathcal{C}}(G)$, we assign the finite group G/U with the discrete topology and, if $U \leq V$ with $V \in \mathcal{N}_{\mathcal{C}}(G)$, then the natural quotient map $G \to G/V$ yields a map $\varphi_{UV} \colon G/U \to G/V$. Furthermore, if $U \leq W \leq V$, then it is clear that $\varphi_{UV} = \varphi_{WV}\varphi_{UW}$. The triple $(G/U, \varphi_{UV}, \mathcal{N}_{\mathcal{C}}(G))$ is an inverse system of non-empty compact Hausdorff groups, so the inverse limit

$$\widehat{G}_{\mathcal{C}} = \varprojlim_{U \in \mathcal{N}_{\mathcal{C}}(G)} G/U$$

is non-empty and it is called the *pro-$\mathcal{C}$ completion* of G. When $\mathcal{C}$ is the class of all finite groups (resp. p-groups), we call it the *profinite* (resp. *pro-p*) completion and we denote it by $\widehat{\cdot}$ (resp. $\widehat{\cdot}_p$).

It is clear that any pro-$\mathcal{C}$ group is also a profinite group. These completions give rise to a plethora of examples of profinite groups, as it allows us to produce one from any abstract group. It does not necessarily mean that the end result will be interesting; for instance, if G is an infinite simple group (meaning, it has no non-trivial normal subgroups), then its

pro-$\mathcal{C}$ completion is trivial for any class of finite groups. We remark also that there exist non-isomorphic abstract groups with isomorphic profinite completions, so some information can be lost in this process.

2.2 The p-adic numbers and $\widehat{\mathbb{Z}}$

Let p be a prime number. We construct an inverse system as follows. If n and m are positive integers with $m \geq n$, then we have the group homomorphism

$$\begin{aligned} \varphi_{m,n} \colon \mathbb{Z}/p^m\mathbb{Z} &\to \mathbb{Z}/p^n\mathbb{Z} \\ x \pmod{p^m} &\mapsto x \pmod{p^n}. \end{aligned}$$

From this, we get an inverse system $(\mathbb{Z}/p^n\mathbb{Z}, \varphi_{m,n}, \mathbb{N})$. The group of p-adic integers is the inverse limit

$$\mathbb{Z}_p = \varprojlim_{n \geq 1} \mathbb{Z}/p^n\mathbb{Z}.$$

We can give very concrete examples of elements which are in the inverse limit and which are not. Let $p = 3$ and consider the following the sequences

$$x = (\ldots, 12 \pmod{81}, 12 \pmod{27}, 3 \pmod{9}, 0)$$

and

$$y = (\ldots, 12 \pmod{81}, 3 \pmod{27}, 3 \pmod{9}, 0).$$

One can check that x is a nested sequence, so $x \in \mathbb{Z}_3$, but $12 \not\equiv 3 \pmod{27}$, hence $y \notin \mathbb{Z}_3$. One observers right away that $\mathbb{Z}_p$ is a pro-p group; in fact, it is the pro-p completion of $\mathbb{Z}$. Furthermore, if we look instead at the profinite completion of $\mathbb{Z}$, then

$$\widehat{\mathbb{Z}} = \prod_p \mathbb{Z}_p$$

and each $\mathbb{Z}_p$ is the Sylow p-subgroup of $\widehat{\mathbb{Z}}$.

Although easy to define, these groups are historically very important. In particular, $\mathbb{Z}_p$ is a *profinite ring*, since each $\mathbb{Z}/p^n\mathbb{Z}$ is a finite ring and the maps we defined are ring homomorphisms. Furthermore, they turn out to be Euclidean domains and their fraction fields are the p-adic rationals $\mathbb{Q}_p$. These, along with $\mathbb{R}$, are all the non-equivalent completions of $\mathbb{Q}$. The p-adic integers have a long tradition as an object of research in various areas of mathematics, such as algebraic number theory and algebraic geometry.

2.3 Galois groups

Another source of examples of profinite groups lies in Galois theory. Specifically, let K be a field and let L be a (not necessarily finite) Galois extension of K. Define

$$\mathcal{N} = \{K \subseteq F \subseteq L : F \text{ is a finite Galois extension}\}.$$

This is a directed partially ordered set with respect to inclusion. To each $F \in \mathcal{N}$, we assign the Galois group $\mathrm{Gal}(F/K)$ and, if $F \subseteq F'$, we define

$$\begin{aligned} \mathrm{res}_{F_2,F_1} \colon \mathrm{Gal}(F'/K) &\to \mathrm{Gal}(F/K) \\ \sigma &\mapsto \sigma_{|_F}. \end{aligned}$$

The triple $(\mathrm{Gal}(F/K), \mathrm{res}_{F',F}, \mathcal{N})$ forms an inverse system and we have

$$\mathrm{Gal}(L/K) = \varprojlim_{F \in \mathcal{N}} \mathrm{Gal}(F/K)\,.$$

Particularly, this shows that every Galois group is a profinite group. Of course, when the extensions are finite, this is what we learn in a first course in Galois theory, along with the theorem of correspondence between intermediate subfields and subgroups of the Galois group. However, to handle infinite extensions (of which there are many important ones, such as the algebraic closure $\overline{\mathbb{Q}}$ of $\mathbb{Q}$), the correct setting is that of profinite groups. In fact, one also has a correspondence theorem here.

Theorem 2.6 (see for instance [RZ10, Theorem 2.11.3]) *Let L/K be a Galois extension of fields with Galois group $G = \mathrm{Gal}(L/K)$. Denote by $\mathcal{I}(L/K)$ the set of intermediate fields $K \subseteq F \subseteq L$ and by $\mathcal{S}(G)$ the set of closed subgroups of G. Then there is a bijection*

$$\Phi \colon \mathcal{S}(G) \to \mathcal{I}(L/K)$$

given by $\Phi(H) = \{x \in L : \sigma(x) = x \text{ for every } x \in H\}$ which reverses inclusions.

Curiously, for a profinite group G, there exists a Galois extension of fields L/K such that $G \cong \mathrm{Gal}(L/K)$ ([RZ10, Theorem 2.11.5]) so every profinite group is also a Galois group.

2.4 p-adic Lie groups

We will be less detailed with the following example, as its setup is somewhat similar to the p-adic integers one. Let p be a prime and consider

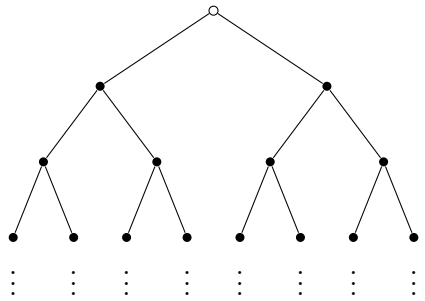
$$\mathrm{SL}_2(\mathbb{Z}_p) = \varprojlim_{n \geq 1} \mathrm{SL}_2(\mathbb{Z}/p^n\mathbb{Z})\,,$$

where the maps are obtained by looking at the entries of the matrix modulo a given power of p. The significance of this example lies in the fact this is a *p-adic analytic group*, that is, a profinite group with a $\mathbb{Q}_p$-analytic structure. In other words, it is a manifold over $\mathbb{Q}_p$ and the group operations are not only continuous, but analytic.

This situation mirrors, in many ways, what happens with real or complex Lie groups: one can associate a particular kind of Lie algebra to it and use it to study the group (and vice-versa). Furthermore, the presence of this analytic structure makes it so that p-adic analytic groups are amenable to analytic tools, something you do not always get for a general profinite group.

2.5 Automorphism groups of rooted trees

Finally, we give one example of a combinatorial nature. Consider the binary rooted tree T as follows.



The root is at level 0 (the white vertex), the next two vertices are at level 1, and so on. Denote by T_n the finite tree obtained by deleting everything below level n. An automorphism of T is a bijection $T \to T$ that sends vertices to vertices, preserves edge incidence and fixes the root vertex. From this, one can see then that $\mathrm{Aut}(T)$ sends T_n to T_n for every n and, intuitively, any automorphism of T can be understood by checking what it does at each truncated finite tree. Thus, we have maps

$$\begin{aligned} \mathrm{Aut}(T_{n+1}) &\to \mathrm{Aut}(T_n) \\ \sigma &\mapsto \sigma_{|_{T_n}} \end{aligned}$$

for each n. This reasoning suggests

$$\mathrm{Aut}(T) = \varprojlim_{n \geq 1} \mathrm{Aut}(T_n);$$

and that is indeed the case. The closed subgroups of $\mathrm{Aut}(T)$ have been largely used to show the existence of profinite groups with many exotic properties; in fact, profinite *just-infinite groups* (groups in which every closed normal subgroup has finite index) are either hereditarily just-infinite or branch; the latter show up as closed subgroups of automorphism groups of rooted trees.

3 Probability in group theory

The interplay between groups and probability has, at this point, quite a long history, with one early example being the results of Erdös–Turan on the statistics of symmetric groups. Since then, the area of asymptotic group theory has both grown as its own subject and been used to solve various problems in other areas of group theory.

A motivating example for my research is the following. For now, let G be a finite group. A very desirable property for G is to be abelian, though of course that is not always the case. It is then natural ask: if not all elements commute, what can we say when a *large*

proportion of them do? This question can be made rigorous when reframed in probabilistic terms. The *commuting probability* of G is defined as

$$\Pr(G) = \frac{\{(x,y) \in G \times G : [x,y] = 1\}}{|G|^2}\,.$$

In other words, we are counting, with uniform probability, the couples of elements which commute. It will be useful to reword this as

$$\Pr(G) = \frac{\{(x,y) \in G \times G : \langle x,y\rangle \text{ is abelian}\}}{|G|^2}\,.$$

Many results can be proven with regards to the commuting probability, but one we would like to (informally) highlight is due to P. M. Neumann: if $\Pr(G) \geq \varepsilon > 0$, then there exists a subgroup H of index bounded by a function of ε with derived subgroup $[H,H]$ of order bounded by a function of ε. This means roughly that a group with large commuting probability contains a 'big' subgroup (i.e., of small index) which is 'close to being abelian' (i.e., it has small derived subgroup).

When dealing with profinite groups, the Haar measure comes into play and we can ask ourselves similarly flavoured questions, and this is the broad theme in which my work is situated.

Before continuing, we recall some classical definitions in group theory and then we define their profinite analogues.

Definition 3.1 Let G be an abstract group. We define

$$G^{(1)} = [G,G] = \gamma_1(G)$$

and, for $n > 1$,

$$G^{(n)} = [G^{(n-1)}, G^{(n-1)}], \quad \gamma_n(G) = [\gamma_{n-1}(G), G]\,.$$

The $G^{(n)}$ and the $\gamma_n(G)$ are the *derived series* and the *lower central series* of G, respectively. We say that G is solvable (resp. nilpotent) if there exists n such that $G^{(n)} = 1$ (resp. $\gamma_n(G) = 1$).

One can think of solvability and nilpotency as 'generalisations' of being abelian. Although one can speak of solvability and nilpotency in the profinite setting, there are correct topological analogues of these conditions.

Definition 3.1 Let G be a profinite group. We define

$$G^{(1)} = \overline{[G,G]} = \gamma_1(G)$$

and, for $n > 1$,

$$G^{(n)} = \overline{[G^{(n-1)}, G^{(n-1)}]}, \quad \gamma_n(G) = \overline{[\gamma_{n-1}(G), G]}\,.$$

The $G^{(n)}$ and the $\gamma_n(G)$ are the *derived series* and the *lower central series* of G, respectively. We say that G is *prosolvable* (resp. *pronilpotent*) if $\cap_{n\geq 1} G^{(n)} = 1$ (resp. $\cap_{n\geq 1}\gamma_n(G) = 1$).

Equivalently, a profinite group is prosolvable (resp. pronilpotent) if it is the inverse limit of solvable groups (resp. nilpotent groups). This is a natural generalisation, as one would want the derived and lower central series to consist of profinite groups, hence why one takes the closure at each step.

We will say that a profinite group G is *virtually* $\mathcal{P}$ if it contains a subgroup of finite-index which satisfies property $\mathcal{P}$. Consider the following (measurable) subset

$$\Omega_{\mathcal{S}}(G,G) = \{(x,y) \in G \times G : \overline{\langle x,y \rangle} \text{ is prosolvable}\}$$

and define analogously $\Omega_{\mathcal{N}}(G,G)$ and $\Omega_p(G,G)$ with pronilpotent and pro-p, respectively, instead of prosolvable. Notice how this mirrors the writing of the commuting probability in terms of elements that generate abelian subgroups. Wilson proved the following.

Theorem 3.3 ([Wil08]) *Let G be a profinite group. If $\mu_{G^2}(\Omega_{\mathcal{S}}(G,G)) > 0$ (respectively $\mu_{G^2}(\Omega_{\mathcal{N}}(G,G)) > 0$), then G is virtually prosolvable (resp. pronilpotent).*

These are profinite analogues of Neumann's result mentioned above, but considering prosolvability and pronilpotency instead of commutativity. With Lucchini, we have shown a similar result for $\Omega_p(G,G)$.

Theorem 3.4 ([LO25, Theorem 6]) *Let G be a profinite group and let p be a prime. If $\mu_{G^2}(\Omega_p(G,G)) > 0$, then G is virtually pro-p.*

More than a decade later, inspired by Wilson's result, Detomi–Lucchini–Morigi–Shumyatsky considered instead the following (measurable) subsets. For $g \in G$, define

$$\Omega_{\mathcal{S}}(g,G) = \{x \in G : \overline{\langle g,x \rangle} \text{ is prosolvable}\}$$

and $\Omega_{\mathcal{N}}(g,G)$ and $\Omega_p(g,G)$ analogously.

Theorem 3.5 ([DLMS23, Theorems 2-3]) *Let G be a profinite group. If $\mu_G(\Omega_{\mathcal{S}}(g,G)) > 0$ (resp. $\mu_G(\Omega_{\mathcal{N}}(g,G)) > 0$) for each $g \in G$, then G is virtually prosolvable (resp. pronilpotent).*

Again, with Lucchini, we proved the pro-p analogue of this result.

Theorem 3.6 ([LO25, Theorem 6]) *Let G be a profinite group and let p be a prime. If $\mu_G(\Omega_p(g,G)) > 0$ for each p-element $g \in G$, then G is virtually pro-p.*

The common theme of these results is that, if G has 'many' subgroups with a (reasonable) property $\mathcal{P}$, then G in fact has a finite-index subgroup with property $\mathcal{P}$.

We mention, however, that this is only one facet of how probability might be used in the study of profinite groups. Much work has been done in other directions and using completely different tools. In the results mentioned above, finite group theory proved essential in all of the results, many of which are consequences of statements of finite groups.

References

[DLMS23] Eloisa Detomi, Andrea Lucchini, Marta Morigi, and Pavel Shumyatsky, *Probabilistic properties of profinite groups*. 2023.

[LO25] Andrea Lucchini and Nowras Otmen, *p-elements in profinite groups*. Monatsh. Math., 206(4): 933–948, 2025.

[RZ10] L. Ribes and P. Zalesskii, "Profinite Groups". Springer-Verlag, Heidelberg, 2010.

[Wat72] W.C. Waterhouse, *An empty inverse limit*. Proceedings of the American Mathematical Society, 36 no. 2: 618, 1972.

[Wil08] John S. Wilson, *The probability of generating a nilpotent subgroup of a finite group*. Bull. Lond. Math. Soc., 40(4): 568–580, 2008.

In Search of Tilting Representations

Anastasios Slaftsos (*)

Abstract. The representation theory of quivers (i.e., oriented graphs encoding algebraic data) consists a central pillar in the study of finite-dimensional algebras. Beyond their combinatorial nature, quivers provide concrete models for more abstract settings, including module categories and their derived counterparts. A fundamental problem in this area is the classification and comparison of representations of a fixed quiver, and tilting theory provides a key mechanism for approaching this problem. In these notes, we introduce tilting theory through explicit examples, beginning with classical tilting modules over finite-dimensional algebras and progressing to more general triangulated settings that extend beyond the classical derived category framework.

1 Quivers and Representations

A **quiver** Q is an oriented graph, given by the data $Q = (Q_0, Q_1, s, t)$, where Q_0 denotes the set of vertices, Q_1 the set of arrows, $s\colon Q_1 \to Q_0$ is the source map, and $t\colon Q_1 \to Q_0$ is the target map.

Example 1.1 The following quivers are going to be running examples throughout these notes.

(a) Consider the quiver $1 \xrightarrow{a} 2 \xrightarrow{b} 3$, whose set of vertices is given by $Q_0 = \{1, 2, 3\}$ and the set of arrows by $Q_1 = \{a, b\}$. In this case, for example, the source of the arrow $a\colon 1 \to 2$ is $s(a) = 1$, whereas its target is $t(a) = 2$. We will denote this quiver by $\mathbb{A}_3$ and call it the **Dynkin quiver of type** A_3.

(b) For the quiver $Q =$  , one has that the set of vertices is the singleton $Q_0 = \{\bullet\}$ and the set of arrows $Q_1 = \{x\}$. We call Q the **Jordan quiver**.

(c) Quivers are not necessarily always finite, in the sense that the set of vertices and the set of arrows can be infinite. For example, one can consider the following quiver

$$Q = \cdots \longrightarrow -2 \xrightarrow{\partial} -1 \xrightarrow{\partial} 0 \xrightarrow{\partial} 1 \xrightarrow{\partial} 2 \xrightarrow{\partial} \cdots ,$$

(*)Ph.D. course, Università di Padova, Dip. Matematica, via Trieste 63, I-35121 Padova, Italy. E-mail: slaftsos@math.unipd.it . Seminar held on 12 March 2026.

whose vertices are indexed by the integer numbers, i.e. $Q_0 = \mathbb{Z}$. Moreover, one can also assume certain **relations** between the arrows. In this case, one can ask for instance that the composition of two consecutive arrows vanishes, that is, $\partial \circ \partial = 0$.

For what follows, we are going to consider $\Bbbk$ to be an algebraically closed field, for example, $\Bbbk = \mathbb{C}$.

Definition 1.2 For a quiver Q, one defines the **path algebra** $A = \Bbbk Q$ of Q, as the $\Bbbk$-vector space, such that

- The basis of this vector space is given by all possible *paths* in Q;
- The unit of the algebra consist of the sum of all the *lazy paths*, that is, paths starting at a vertex and ending at the same vertex without doing anything in between;
- Multiplication is given by concatenation of paths.

Example 1.3 One can explicitly compute the path algebras of the quivers in Example 1.1.

(1) For the Dynkin quiver $\mathbb{A}_3 = 1 \xrightarrow{a} 2 \xrightarrow{b} 3$, the path algebra $\Bbbk\mathbb{A}_3$ is a six-dimensional $\Bbbk$-vector space with basis $\{e_1, e_2, e_3, a, b, ba\}$, where e_i, $i = 1, 2, 3$, denotes the lazy paths in vertices $1, 2$ and 3, respectively, and ba denotes the path of length 2 from the vertex 1 to the vertex 3. The path algebra $\Bbbk\mathbb{A}_3$ in this case is isomorphic (as a ring) to the ring of lower triangular matrices

$$T_3 = \begin{pmatrix} \Bbbk & 0 & 0 \\ \Bbbk & \Bbbk & 0 \\ \Bbbk & \Bbbk & \Bbbk \end{pmatrix}.$$

(2) The path algebra of the quiver $Q = \bullet \circlearrowright x$ is an infinite-dimensional vector space with basis $\{e, x^n : n \in \mathbb{Z}_{\geq 0}\}$, where e is the lazy path at the single vertex of Q. The path algebra $\Bbbk Q$ is isomorphic (as a ring) to the polynomial ring $\Bbbk[x]$.

(3) The example of the quiver $Q = \cdots \longrightarrow -2 \xrightarrow{\partial} -1 \xrightarrow{\partial} 0 \xrightarrow{\partial} 1 \xrightarrow{\partial} 2 \xrightarrow{\partial} \cdots$, will be revisited later, since in this case, one cannot define a ring structure for the path algebra of Q.

Definition 1.4 A **representation** of a quiver Q is a pair $((V_i)_{i \in Q_0}, (f_a)_{a \in Q_1})$, such that each V_i, $i \in Q_0$ is a $\Bbbk$-vector space, and for any arrow $a : i \to j$ in Q_1, the map $f_a : V_i \to V_j$ is linear.

A **morphism of representations** $\varphi\colon((V_i)_{i\in Q_0},(f_a)_{a\in Q_1})\to((W_i)_{i\in Q_0},(g_a)_{a\in Q_1})$ is a family of linear maps $(\varphi_i\colon V_i\to W_i)_{i\in Q_0}$, such that, for any arrow $a\colon i\to j$ in Q_1, the square

$$\begin{array}{ccc} i & & V_i & \xrightarrow{\varphi_i} & W_i \\ {\scriptstyle a}\downarrow & & {\scriptstyle f_a}\downarrow & & \downarrow{\scriptstyle g_a} \\ j & & V_j & \xrightarrow{\varphi_j} & W_j \end{array},$$

commutes, that is, $g_a\varphi_i=\varphi_j f_a$. A morphism of representations is said to be an **isomorphism**, if φ_i is an isomorphism of vector spaces for every $i\in Q_0$.

Notice that representations of a quiver forms a category (see [ML98]). The objects of this category are quiver representations and the morphisms are morphisms of representations. We will denote by $\mathrm{Rep}(Q)$ the category of all representations of Q and by $\mathrm{rep}(Q)$ the category of the finite-dimensional representations of Q.

A guiding problem in representation theory of quivers is the classification of all representations of a given quiver, up to isomorphism. This can often be very difficult (if not impossible), however, one can reduce this problem, to the classification of "well-behaved" representations which act as the *building blocks* of the category $\mathrm{rep}(Q)$ of finite-dimensional representations.

Definition 1.5 A representation M of a quiver Q is said to be **indecomposable**, if it is not isomorphic to a direct sum of two other (non-zero) representations.

Example 1.6 Consider the quiver $\mathbb{A}_3=1\xrightarrow{a}2\xrightarrow{b}3$ and the representations

$$M=\Bbbk^2\xrightarrow{\begin{pmatrix}1&0\end{pmatrix}}\Bbbk\xrightarrow{0}0 \quad\text{and}\quad N=\Bbbk\xrightarrow{\begin{pmatrix}1\\-1\\1\end{pmatrix}}\Bbbk^3\xrightarrow{\begin{pmatrix}1&1&0\\0&1&1\end{pmatrix}}\Bbbk^2$$

A morphism $\varphi\colon M\to N$ is given by

$$\begin{array}{ccc} M & & \Bbbk^2 & \xrightarrow{\begin{pmatrix}1&0\end{pmatrix}} & \Bbbk & \xrightarrow{0} & 0 \\ {\scriptstyle\varphi}\downarrow & & \begin{pmatrix}1&0\end{pmatrix}\downarrow & & \downarrow\begin{pmatrix}1\\-1\\1\end{pmatrix} & & \downarrow 0 \\ N & & \Bbbk & \xrightarrow[\begin{pmatrix}1\\-1\\1\end{pmatrix}]{} & \Bbbk^3 & \xrightarrow[\begin{pmatrix}1&1&0\\0&1&1\end{pmatrix}]{} & \Bbbk \end{array}$$

Notice also that the representation M is **decomposable** since it can be written as the direct sum of two other *indecomposable* representations, as follows:

$$M \quad = \quad \Bbbk\xrightarrow{0}0\longrightarrow 0 \quad \oplus \quad \Bbbk\xrightarrow{\mathrm{id}}\Bbbk\longrightarrow 0\ .$$

In fact this is not a strange phenomenon, as one can see in the following Theorem.

Theorem 1.7 (Krull-Remak-Schmidt) *Let Q be a quiver and M a finite-dimensional representation of Q. Then, M decomposes uniquely as a direct sum of indecomposable representations.*

As we mentioned above, the indecomposable representations behave like building blocks for all finite dimensional representations of a given quiver. In particular, they encode all the information of the category of finite dimensional representations $\mathrm{rep}(Q)$. A fundamental concept that illustrates perfectly this phenomenon is that of the Auslander-Reiten quiver.

Definition 1.8 The **Auslander-Reiten quiver** of a quiver Q is a quiver $\Gamma(Q)$ defined as follows.

- The vertices are finite-dimensional indecomposable representations of Q;
- The arrows are irreducible morphisms between finite-dimensional indecomposable representations.

Example 1.9 We construct the Auslander-Reiten quiver of the quiver $\mathbb{A}_3 = 1 \xrightarrow{a} 2 \xrightarrow{b} 3$. We begin by finding all the indecomposable finite-dimensional representations of $\Bbbk\mathbb{A}_3$. One can check that the indecomposable representations are in fact six, and are given by:

- $(1\,1\,1) = \Bbbk \xrightarrow{1} \Bbbk \xrightarrow{1} \Bbbk$, also denote by P_1 and I_3;
- $(0\,1\,1) = 0 \to \Bbbk \xrightarrow{1} \Bbbk$, also denoted by P_2;
- $(0\,0\,1) = 0 \to 0 \to \Bbbk$, also denoted by P_3 and S_3;
- $(1\,0\,0) = \Bbbk \to 0 \to 0$, also denoted by I_1 and S_1;
- $(1\,1\,0) = \Bbbk \xrightarrow{1} \Bbbk \to 0$, also denoted by I_2;
- $(0\,1\,0) = 0 \to \Bbbk \to 0$, also denoted by S_2.

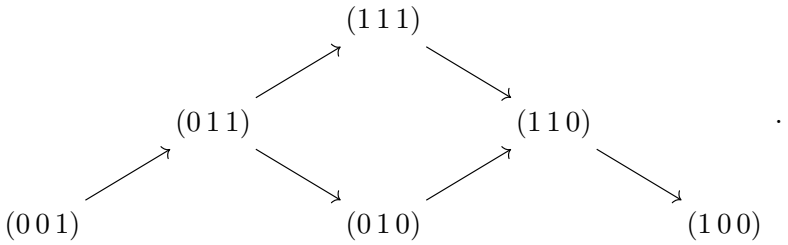
The representations P_1, P_2 and P_3 are called indecomposable **projective** representations, their duals I_1, I_2 and I_3 are called indecomposable **injective** representations, and finally, the representations S_1, S_2 and S_3 are called **simple** representations.

One can also explicitly compute all the irreducible morphisms between the indecomposable representations. The computations are left to the reader, however we do include a list of all irreducible morphisms. In particular, there is:

- A monomorphism $S_3 \to P_2$;
- A monomorphism $P_2 \to P_3$;
- A monomorphism $S_2 \to I_2$;

- An epimorphism $P_2 \to S_2$;
- An epimorphism $P_1 \to I_2$;
- An epimorphism $I_2 \to S_1$.

Arranging the above listed information into an oriented graph, yields the Auslander-Reiten quiver of $\mathbb{A}_3$. Diagrammatically, the Auslander-Reiten quiver is given by the following graph



We conclude this first introductory section with a fundamental result in representation theory of algebras that illustrates the importance of quivers in modern algebra.

Theorem 1.10 *Let Λ be a finite-dimensional algebra. Then, there exists a quiver Q (possibly with relations), such that, the category of Λ-modules is equivalent to the category of representations of Q. In symbols,*

$$\operatorname{Mod}\Lambda \xrightarrow{\cong} \operatorname{Rep}(Q)$$

is an equivalence of categories.

2 Tilting Theory

In order to develop tilting theory, we first need to establish the appropriate setup for this theory, which turns out to be the one of the derived category.

Let R be a ring and let $\operatorname{Mod} R$ denote the category of (right) R-modules. The sequence of R-modules

$$(\#1) \qquad X = \cdots \longrightarrow X^{n-1} \xrightarrow{d^{n-1}} X^n \xrightarrow{d^n} X^{n+1} \longrightarrow \cdots,$$

where for every $n \in \mathbb{Z}$, one has $d^n d^{n-1} = 0$, is called a (cochain) **complex of R-modules**. A morphism φ between two complexes X, Y is called a **chain map** and is given by a family $\{\phi_i\}_{i\in\mathbb{Z}}$ of R-linear maps

$$\begin{array}{ccccccccccc}
X & & \cdots & \longrightarrow & X^{n-1} & \xrightarrow{d_X^{n-1}} & X^n & \xrightarrow{d_X^n} & X^{n+1} & \longrightarrow & \cdots \\
{\scriptstyle\varphi}\downarrow & & & & \downarrow{\scriptstyle\varphi^{n-1}} & & \downarrow{\scriptstyle\varphi^n} & & \downarrow{\scriptstyle\varphi^{n+1}} & & \\
Y & & \cdots & \longrightarrow & Y^{n-1} & \xrightarrow[d_Y^{n-1}]{} & Y^n & \xrightarrow[d_Y^n]{} & Y^{n+1} & \longrightarrow & \cdots
\end{array}$$

such that, every square commutes, that is, $d_Y^n \varphi^n = \varphi^{n+1} d_X^n$ for every $n \in \mathbb{Z}$. We denote the category of complexes of R-modules by $\mathbf{Ch}(R)$. We call a complex X **acyclic**, if the sequence (#1) is exact, that is $\operatorname{Im} d^{n-1} = \operatorname{Ker} d^n$, for every $n \in \mathbb{Z}$. Notice that the relation $d^n d^{n-1} = 0$, implies the inclusion of the image into the kernel. For any $n \in \mathbb{Z}$, we define the **cohomology functor**:

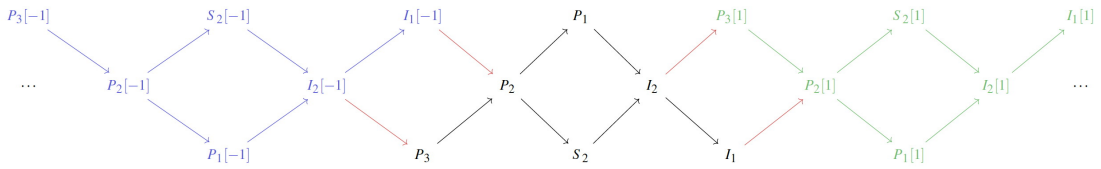
$$H^n(-) \colon \mathbf{Ch}(R) \longrightarrow \mathsf{Ab},$$

that maps a complex (X, d) to the quotient abelian group $H^n(X) = \operatorname{Ker} d^n / \operatorname{Im} d^{n-1}$. In essence, the cohomology functor measures how far away is a complex from being exact. A chain map $\varphi \colon X \longrightarrow Y$ is said to be a **quasi-isomorphism**, if the induced map in the cohomology level is an isomorphism of abelian groups, that is $H^n(f) \colon H^n(X) \xrightarrow{\sim} H^n(Y)$ for every $n \in \mathbb{Z}$. We are particularly interested in formally inverting the quasi-isomorphisms into proper isomorphisms of complexes. This technique is called **localisation** with respect to the class of quasi-isomorphisms.

The **derived category** of the ring R is defined as the localisation of the category of complexes of R-modules with respect to the class of quasi-isomorphisms and it is denoted by $\mathbf{D}(R)$. Note that, in the derived category, quasi-isomorphisms become isomorphisms, meaning that complexes with the same cohomology are identified. In particular, every acyclic complex is isomorphic to zero in $\mathbf{D}(R)$.

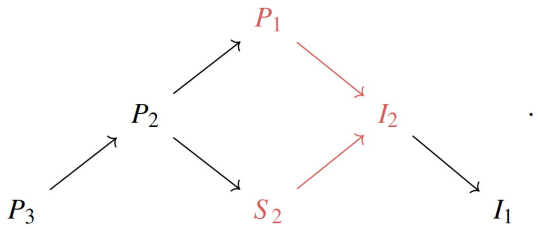
Derived categories provide the correct framework to compare various homological invariants of rings such as K-theory, Hochschild cohomology, (finiteness of) global dimension, and many others. More concretely, if two rings are isomorphic, then all of their invariants are isomorphic as well. In fact, as shown by Morita in [Mor58], it suffices that the rings have equivalent module categories. However, both instances are quite restrictive, whereas, when one works in the setup of the derived category, the necessary conditions to compare invariants of rings are relaxed, meaning that there is no requirement of equivalent representations.

Example 2.1 The (bounded) derived category of the path algebra $A = \Bbbk\mathbb{A}_3$ of the quiver Dynkin quiver $\mathbb{A}_3$, is given by glueing $\mathbb{Z}$-copies of the Auslander-Reiten quiver. In particular, one can draw the Auslander-Reiten quiver of the bounded derived category as follows:

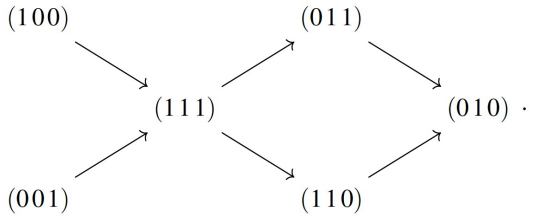


where we highlight with blue the copy of the Auslander-Reiten quiver of $\mathbb{A}_3$ in degree -1 (and all of the copies in negative degrees that are omitted in the graph above) and with green the copy in degree 1(and all of the copies in positive degrees that are omitted in the graph above). Note that the red arrows represent extensions between stalk complexes in the bounded derived category.

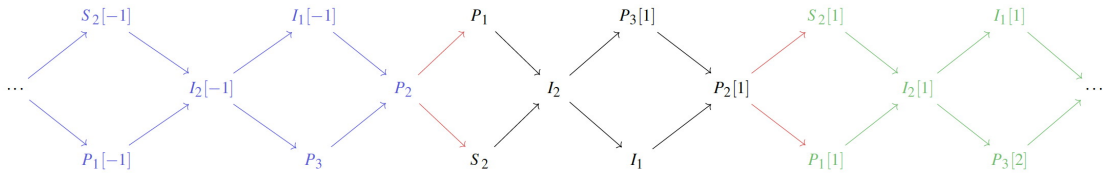
Consider now the representation $T = P_1 \oplus I_2 \oplus S_2$ in $\mathrm{Rep}(\mathbb{A}_3)$. We want to compute the endomorphism ring of T in $\mathrm{Rep}(\mathbb{A}_3)$. Recall that the Auslander-Reiten quiver of $\mathbb{A}_3$ contains all the data that one needs in order to compute morphisms between finite-dimensional representations. In particular, in the diagram below we highlight with red the indecomposable direct summands of T and the morphisms between them



It turns out that the endomorphism ring $B = \mathrm{End}_A(T)$ can be computed explicitly as the opposite quiver of the highlighted red quiver, that is the quiver with the same vertices and arrows going to the opposite direction. That is, B is the path algebra of the quiver $Q' = 1 \longleftarrow 2 \longrightarrow 3$. One can compute the Auslander-Reiten quiver of Q' as in Example 1.9, and obtain the diagram below



Similarly, we compute the Auslander-Reiten quiver of the bounded derived category of B and draw it as follows:



where, as before, we highlight with blue the copy of the Auslander-Reiten quiver of Q' in degree -1 and with green the copy in degree 1. Notice that the Auslander-Reiten quivers of the bounded derived categories of A and B coincide, implying that the two bounded derived categories are equivalent.

The equivalence of derived categories in the Example above, is not a coincidence. In fact, representations that share the same properties as the representation T from Example 2.1, are called tilting representations and play a fundamental role in derived equivalences.

Definition 2.2 A finite-dimensional representation T of a quiver Q is called a **tilting representation** if it has no self-extensions, that is $\mathrm{Ext}^1(T,T) = 0$, and the number of indecomposable direct summands of T equals the number of vertices of Q.

Theorem 2.3 ([Hap88]) *If T is a tilting representation of the path algebra $A = \Bbbk Q$ of a quiver Q, then there is an equivalence*

$$\mathbf{D}^b(A) \xrightarrow{\sim} \mathbf{D}^b(\mathrm{End}_A(T))$$

between the bounded derived category of A and the bounded derived category of the endomorphism ring of T.

3 Q-shaped Homological Algebra

In this final section we are going to briefly overview recent developments in homological algebra. In particular, as we mentioned in the previous section, the correct framework to develop homological algebra turns out to be derived categories. As we will explain below, one can even replace derived categories of complexes, with other categories that share similar properties, by replacing the underlying quiver of complexes, with various suitably nice quivers.

Consider the quiver $Q = \cdots \longrightarrow -1 \xrightarrow{\partial} 0 \xrightarrow{\partial} 1 \longrightarrow \cdots$ with the relation: every two consecutive arrows compose to zero, that is $\partial^2 = 0$. It is clear that the category of representations of Q coincides with the category of complexes of $\Bbbk$-vector spaces. In symbols, one has

$$\mathrm{Rep}(Q) = \mathbf{Ch}(\Bbbk) =: \mathrm{Mod}\, Q.$$

The main idea is to replace the underlying quiver Q with different "shapes" and do homological algebra without complexes, unifying the study of various categories into a single theory. In the following example we list some of these categories and admissible "shapes".

Example 3.1 (1): Consider Q to be the following quiver

$$Q = \qquad \cdots \longrightarrow \bullet \xrightarrow{\delta} \bullet \xrightarrow{\delta} \bullet \xrightarrow{\delta} \bullet \longrightarrow \cdots, \qquad \delta^N = 0$$

where the vertices are indexed by the integers and the composition of N-consecutive arrows is zero, for some $N \geq 2$. The category $\mathrm{Mod}\, Q$ is equivalent to the category of N-complexes of $\Bbbk$-vector spaces (see for example [IKM17]).
(2): Consider the cyclic quiver Q^m given by a set of m vertices and m-arrows between them such as any two consecutive arrows compose to zero (see Figure 1).

Holm and Jørgensen in [HJ22] defined for "nice" quivers Q the notion of the Q-shaped derived category, which we will denote by $\mathbf{D}_Q$. The main techniques of this construction are based on the same ideas for constructing the derived category from the category of complexes. Namely, one needs to define acyclic objects and a suitable cohomology functor that detects them, identify objects with the same cohomology via some notion of a *weak-equivalence* and then formally invert these weak-equivalences into proper isomorphisms.

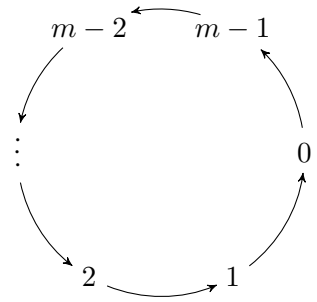


Figure 1: Cyclic quiver with m vertices and relations: every two consecutive arrows compose to zero.

Notice that the objects in the Q-shaped derived category are *not* complexes, but quiver representations (modulo some homotopy). Categories like the derived category of a ring or the Q-shaped derived category are called triangulated.
In this case, the category $\operatorname{Mod} Q$ coincides with the category of m-periodic complexes.

Definition 3.2 A **triangulated category** is an additive category $\mathscr{T}$ endowed with an auto-equivalence functor $[1]\colon \mathscr{T} \longrightarrow \mathscr{T}$, called the suspension or shift, and a class of triangles of the form $X \to Y \to Z \to X[1]$, satisfying certain axioms.

One can define tilting objects in every triangulated category, extending the notion of tilting representations in the derived category of path algebras.

Definition 3.3 An object T in a triangulated category $\mathscr{T}$ is called **tilting** if,

(1) it has no self-extensions, that is, $\operatorname{Hom} \mathscr{T}TT[n] = 0$, for all $n \in \mathbb{Z}$, and

(2) it generates $\mathscr{T}$, that is, if $\operatorname{Hom} \mathscr{T}TX[n] = 0$ for all $n \in \mathbb{Z}$ then $X = 0$.

Naturally arises the question, whether every triangulated category admits a tilting object. The answer to this is negative, since there exist triangulated categories with no non-zero tilting objects. We reformulate this question in the context of Q-shaped derived categories.

Question: Does every Q-shaped derived category admit a tilting object?

In [Sla26] we approach this problem by studying t-structures in such categories. Below we give the definition of a t-structure and the most classical example, that of the standard t-structure of the derived category of a ring, adapted in the case of the bounded derived category of $\mathbb{A}_3$.

Definition 3.4 A pair $(\mathcal{U}, \mathcal{V})$ of subcategories in a triangulated category $\mathscr{T}$ is called a t-structure if it satisfies the following properties.

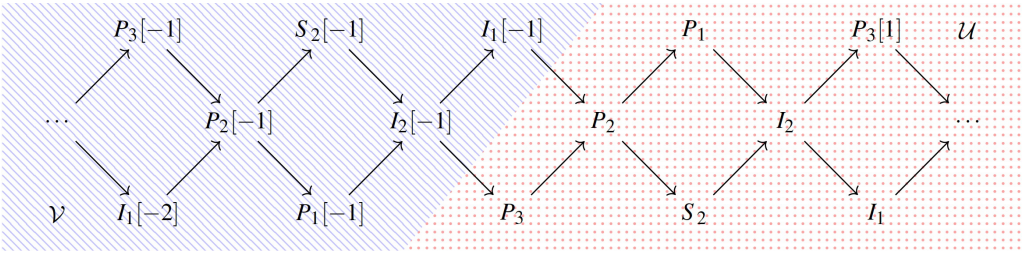
(1) $\operatorname{Hom} \mathscr{T}UV = 0$ for every $U \in \mathcal{U}$ and $V \in \mathcal{V}$;

(2) $\mathcal{U}[1] \subseteq \mathcal{U}$ (or $\mathcal{V}[-1] \subseteq \mathcal{V}$);

(3) Every object $T \in \mathscr{T}$ fits into a triangle $U \to T \to V \to U[1]$, with $U \in \mathcal{U}$ and $V \in \mathcal{V}$.

Theorem 3.5 ([BBD82]) *The* ***heart*** $\mathcal{H} = \mathcal{U} \cap \mathcal{V}[1]$ *of a t-structure* $(\mathcal{U}, \mathcal{V})$ *in a triangulated category* $\mathscr{T}$*, is an abelian category.*

Example 3.6 Let $\mathscr{T}$ be the bounded derived category of $\mathbb{A}_3$, that is $\mathscr{T} = \mathbf{D}^b(\Bbbk\mathbb{A}_3)$ and consider the subcategories $\mathcal{U}$ and $\mathcal{V}$ as in the following diagram



where we distinguish $\mathcal{V}$ in the dashed blue area and $\mathcal{U}$ in the dotted red area. To compute the heart, one first has to apply the functor [1] to $\mathcal{V}$ and then compute the intersection $\mathcal{U} \cap \mathcal{V}[1]$. In this case, the resulting abelian category is nothing more than $\mathrm{Rep}(\mathbb{A}_3)$. One can easily visualise this in the diagram above, by observing that the underlying quiver of the heart is the Auslander-Reiten quiver of $\mathbb{A}_3$.

Our strategy on checking whether a Q-shaped derived category admits a tilting object or not, relies on the following defining property of tilting objects. In the following Proposition, we are not going to rigorously define what properties "nice" t-structures and what "nice" hearts have, since they are outside of the scope of this manuscript. Thus we will use the next Proposition as a black box.

Proposition 3.7 *Every tilting object* T *in a triangulated category* $\mathscr{T}$ *induces a "nice" t-structure with a "nice" heart* $\mathcal{H}$ *such that*

$$\mathscr{T} \xrightarrow{\sim} \mathbf{D}(\mathcal{H}).$$

In order to guarantee the existence of tilting objects in the Q-shaped derived category, we need to impose a combinatorial condition on Q. Notice, that for the construction of the Q-shaped derived category the quiver Q has to satisfy the setup [HJ22, Stp. 2.5]. Note that the existence of a Serre functor in this setup, excludes finite quivers except if they have cycles.

Definition 3.8 Let $\mathcal{L}, \mathcal{R}$ be classes of objects of Q that consist a partition of Q, such that $\mathcal{L}$ is subject to the following conditions:

(1) The class $\mathcal{L}$ is closed under predecessors;

(2) There does *not* exist an infinite sequence $q_1 \to q_2 \to q_3 \to \cdots$ of non-zero morphisms in the pseudo-radical $\mathfrak{r}$ with $q_1, q_2, q_3, \ldots \in \mathcal{L}$.

We call such a partition of Q an **admissible partition**.

Definition 3.9 Let $\mathcal{L}, \mathcal{R}$ be an admissible partition of Q. We define the **boundary** of $\mathcal{L}$ (respectively of $\mathcal{R}$) in the following sense

$$\partial\mathcal{L} := \{q \in \mathcal{L} \,|\, \text{There exists some object } p \in \mathcal{R} \text{ such that } Q(q,p) \neq 0\}$$
$$\partial\mathcal{R} := \{q \in \mathcal{R} \,|\, \text{There exists some object } p \in \mathcal{L} \text{ such that } Q(p,q) \neq 0\}$$

We denote by $\mathrm{Rep}(\partial\mathcal{L})$ the category of quiver representation of the boundary quiver $\partial\mathcal{L}$.

For quivers that satisfy the setup of [HJ22] and have an admissible partition as in Definition 3.8, it is proven in [Sla26] that the Q-shaped derived category admits a silting (and in some cases even a tilting) object, that induces a t-structure whose heart is the category of representations of the boundary quiver. We will not delve deeper in this result, but we will showcase it in the following Examples, to recover some well-known equivalences in the literature.

Example 3.10 Consider Q to be the following quiver

$$Q = \qquad \cdots \longrightarrow \bullet \xrightarrow{\delta} \bullet \xrightarrow{\delta} \bullet \xrightarrow{\delta} \bullet \longrightarrow \cdots, \qquad \delta^N = 0$$

where the vertices are indexed by the integers and the composition of N consecutive arrows is zero, for some $N \geq 2$. It is proven in [HJ22] that the Q-shaped derived category $\mathbf{D}_Q$ is equivalent to the derived category of N-complexes $\mathbf{D}_N(\Bbbk)$ of $\Bbbk$-vector spaces. In this setup, consider the admissible partition $\mathcal{L}, \mathcal{R}$ of Q given by

$$\mathcal{L} = \{q \in \mathbb{Z} \,|\, q \leq 0\} \quad \text{and} \quad \mathcal{R} = \{q \in \mathbb{Z} \,|\, q > 0\}.$$

The boundary $\partial\mathcal{L}$ of $\mathcal{L}$ is highlighted in red in the following graph:

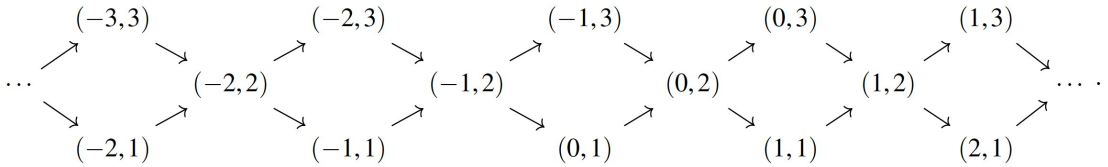
It is straightforward to check that $\mathrm{Rep}(\partial\mathcal{L})$ coincides with the category of representations of the Dynkin quiver $\mathbb{A}_{N-1}$. In this case, the tilting object is given by $T = \bigoplus_{i=1}^{N-1} P_i$, the direct sum of the indecomposable projective representations of $\mathbb{A}_{N-1}$, and one has a triangle equivalence

$$\mathbf{D}_Q(\Bbbk) \xrightarrow{\cong} \mathbf{D}(T_{N-1}(\Bbbk)),$$

where $\mathbf{D}(T_{N-1}(\Bbbk))$ derived category of upper triangular $(N-1) \times (N-1)$ matrices, or equivalently the derived category of the category $\mathrm{Rep}(\mathbb{A}_{N-1})$ of representations of $\mathbb{A}_{N-1}$ (see also [IKM17, Prop. 4.11]).

In the special case that $N = 2$ then $\mathbf{Ch}_N(\Bbbk)$ is simply the (classical) category of complexes of $\Bbbk$-vector spaces $\mathbf{Ch}(\Bbbk)$. The above construction recovers the standard t-structure in the derived category $\mathbf{D}(\Bbbk)$.

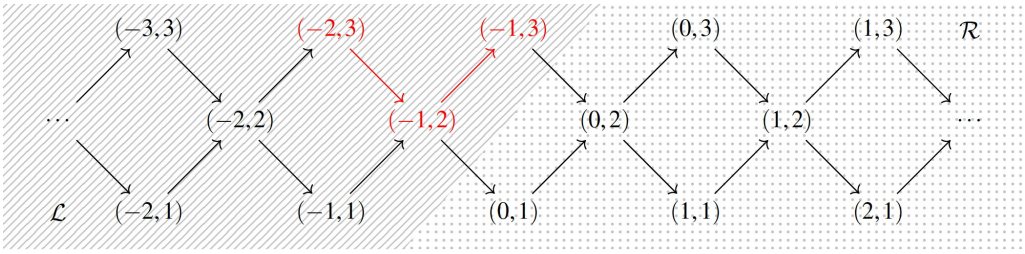
Example 3.11 Consider Q to be given by the repetitive quiver of $\mathbb{A}_3$ modulo the mesh relations, or schematically by the following graph



We define an admissible partition $\mathcal{L}, \mathcal{R}$ of Q given by

$$\mathcal{L} = \{(i,j) \in Q \mid i < 0\} \quad \text{and} \quad \mathcal{R} = \{(i,j) \in Q \mid i \geq 0\}.$$

In the following graph we distinguish $\mathcal{L}$ within the striped area, $\mathcal{R}$ within the dotted area, and we highlight the boundary $\partial\mathcal{L}$ of $\mathcal{L}$ with red.



It is straight forward to check that $\mathrm{Rep}(\partial\mathcal{L})$ coincides with the category of representations of the quiver

$$\Delta = \qquad \bullet \xrightarrow[a]{} \bullet \xrightarrow[b]{} \bullet, \qquad ba = 0 \ .$$

The tilting object is given by the direct sum of the indecomposable projective representations, that is:

$$T = \qquad \Big(0 \longrightarrow 0 \longrightarrow \Bbbk\Big) \qquad \oplus \qquad \Big(0 \longrightarrow \Bbbk \xrightarrow{1} \Bbbk\Big) \qquad \oplus \qquad \Big(\Bbbk \xrightarrow{1} \Bbbk \longrightarrow 0\Big) \ .$$

and induces a triangle equivalence,

$$\mathbf{D}_Q \xrightarrow{\cong} \mathbf{D}(\mathrm{Rep}(\Delta)) \xrightarrow{\cong} \mathbf{D}(\Bbbk\mathbb{A}_3),$$

which recovers the already known result in [GHJS24].

References

[BBD82] A.A. Beilinson, J. Bernstein, and P. Deligne, *Faisceaux pervers*. In Analysis and Topology on Singular Spaces, I (Luminy, 1981), volume 100 of Astérisque, pages 5–171. Soc. Math. France, Paris, 1982.

[GHJS24] Sira Gratz, Henrik Holm, Peter Jorgensen, and Greg Stevenson, *Tilting in Q-shaped derived categories*. arXiv preprint arXiv:2411.11412, 2024.

[Hap88] Dieter Happel, "Triangulated categories in the representation theory of finite-dimensional algebras". Volume 119 of London Math. Soc. Lecture Note Ser. Cambridge University Press, Cambridge, 1988.

[HJ22] Henrik Holm and Peter Jørgensen, *The Q-shaped derived category of a ring*. J. Lond. Math. Soc. (2), 106(4): 3263–3316, 2022.

[IKM17] Osamu Iyama, Kiriko Kato, and Jun-ichi Miyachi, *Derived categories of N-complexes*. J. Lond. Math. Soc. (2), 96(3): 687–716, 2017.

[ML98] Saunders Mac Lane, "Categories for the working mathematician". Volume 5 of Grad. Texts in Math. Springer-Verlag, New York, second edition, 1998.

[Mor58] Kiiti Morita, *Duality for modules and its applications to the theory of rings with minimum condition*. Sci. Rep. Tokyo Kyoiku Daigaku Sect. A, 6: 83–142, 1958.

[Sla26] Anastasios Slaftsos, *Silting t-structures in Q-shaped derived categories*. arXiv preprint. arXiv:2606.20447, 2026.

An introduction to infinite-dimensional geometry in continuum mechanics

Francesca Berlinghieri (*)

Abstract. Continuum mechanics studies the motion and deformation of bodies modeled as continuous media. It covers a large number of theories, including ideal, compressible and viscous fluid models, as well as linear and nonlinear elasticity. In these notes, a transition from Newtonian mechanics of discrete systems to the setting of continuum mechanics will be presented through the principle of virtual works. Particular emphasis will be given to the role of the placement (or deformation) map as primary descriptor of the continuum. Such maps naturally belong to functional spaces that are inherently infinite-dimensional. This perspective motivates the study of the geometry of spaces of admissible deformations. As a classical example, we will briefly discuss the approach introduced by Arnold, and later developed by Ebin and Marsden, where the motion of an incompressible perfect fluid is interpreted as a geodesic flow on the (infinite-dimensional) manifold of volume-preserving diffeomorphisms.

1 Introduction

Continuum mechanics provides a mathematical framework for describing the motion and deformation of bodies whose mass is continuously distributed in space. Unlike classical Newtonian mechanics, where the state of a system is determined by the positions of finitely many particles, continuum theories involve infinitely many degrees of freedom and therefore require a different analytical setting. A natural way to bridge these two perspectives is through the principle of virtual works. In the finite-dimensional setting of Newtonian mechanics, this principle is equivalent to the classical equations of motion and provides a formulation that extends naturally to continua. The resulting description highlights the role of the placement map, which associates to each material point its position in the ambient space and serves as the fundamental kinematic descriptor of the motion. This observation suggests that the motion of a continuum may be studied through the geometric properties of suitable (infinite-dimensional) spaces of deformations. The purpose of these notes is to introduce this viewpoint. After reviewing the principle of virtual works and the basic concepts of continuum mechanics, we present a classical example due to Arnold [1]

(*)Ph.D. course, Università di Padova, Dip. Matematica, via Trieste 63, I-35121 Padova, Italy. E-mail: berlingh@math.unipd.it . Seminar held on 26 March 2026.

and later developed by Ebin and Marsden [3], in which the motion of an incompressible perfect fluid is interpreted as a geodesic flow on an infinite-dimensional manifold of volume-preserving diffeomorphisms.

2 Newtonian mechanics and equations of motion

Consider a discrete system of N material points with constant masses $m_1, \dots, m_N$ and position vectors $\boldsymbol{r}_1(t), \dots, \boldsymbol{r}_N(t) \in \mathbb{R}^3$, which describe the position of the material point labeled by $\lambda \in \{1, \dots, N\}$ at time t. The equations of motion of such a system are given by Newton's second law as

$$\begin{aligned} m_1 \ddot{\boldsymbol{r}}_1(t) &= \boldsymbol{f}_1(\boldsymbol{r}_1(t), \dots, \boldsymbol{r}_N(t), \dot{\boldsymbol{r}}_1(t), \dots, \dot{\boldsymbol{r}}_N(t), t), \\ &\vdots \\ m_N \ddot{\boldsymbol{r}}_N(t) &= \boldsymbol{f}_N(\boldsymbol{r}_1(t), \dots, \boldsymbol{r}_N(t), \dot{\boldsymbol{r}}_1(t), \dots, \dot{\boldsymbol{r}}_N(t), t). \end{aligned} \tag{1}$$

where a superimposed dot denotes time differentiation. The force acting on the λ-th particle, denoted by $\boldsymbol{f}_\lambda$, depends on time, positions, velocities and on the label $\lambda \in \{1, \dots, N\}$.

If we now introduce virtual velocities $\boldsymbol{w}_1(t), \dots, \boldsymbol{w}_N(t)$ that, for each time t, represent a generic element of the tangent spaces to $\mathbb{R}^3$ at the points $\boldsymbol{r}_1(t), \dots, \boldsymbol{r}_N(t)$, we can take the scalar product of each equation in (1) with the corresponding virtual velocity, then sum over the point labels and integrate over time within a given interval $[t_i, t_f]$ and obtain

$$\sum_{\lambda=1}^{N} \int_{t_i}^{t_f} (m_\lambda \ddot{\boldsymbol{r}}_\lambda \cdot \boldsymbol{w}_\lambda - \boldsymbol{f}_\lambda \cdot \boldsymbol{w}_\lambda)\, dt = 0.$$

Integrating by parts and assuming that $\boldsymbol{w}_\lambda(t_i) = \boldsymbol{w}_\lambda(t_f) = 0$ for simplicity, we get

$$\sum_{\lambda=1}^{N} \int_{t_i}^{t_f} (-\boldsymbol{p}_\lambda \cdot \dot{\boldsymbol{w}}_\lambda - \boldsymbol{f}_\lambda \cdot \boldsymbol{w}_\lambda)\, dt = 0, \tag{2}$$

where

$$\boldsymbol{p}_\lambda := m_\lambda \dot{\boldsymbol{r}}_\lambda$$

is the linear momentum of the λ-th particle. Both $\boldsymbol{p}_\lambda$ and $\boldsymbol{f}_\lambda$ are, in general, functions of label, time, positions and velocities. Assuming that (2) holds for every choice of smooth virtual velocities that vanish at the initial and final time is an equivalent formulation of Newton's equations (1). This reformulation is known as the *principle of virtual works*. As we shall see, the principle of virtual works can be generalized to the case of a continuum more easily than Newton's equations.

3 Kinematics of continua

In continuum mechanics, matter is modeled as a continuous body rather than as a discrete collection of particles. A fundamental role in defining a continuum model is played by

the set of labels $\Lambda \subset \mathbb{R}^3$, assumed to be open and connected, that is also called *material manifold*. We can interpret $\boldsymbol{\lambda} \in \Lambda$ as the vector of coordinates that identifies the position of a material point in the set of labels, and hence we attribute to $\boldsymbol{\lambda}$ the units of a length. We assume that material points are positioned in a three-dimensional Euclidean space (also called the *ambient space*) and the position coordinates, on a fixed frame of reference, of the point $\boldsymbol{\lambda}$ at time t are given by the *placement* or *deformation map*

$$\begin{aligned} \boldsymbol{\varphi}_t : \Lambda &\to \mathbb{R}^3 \\ \boldsymbol{\lambda} &\mapsto \boldsymbol{x}, \end{aligned}$$

that associates to each material point labeled by $\boldsymbol{\lambda}$ its position $\boldsymbol{x} = \boldsymbol{\varphi}_t(\boldsymbol{\lambda})$ in the ambient space at time t. This map is assumed to be, for every time t, a diffeomorphism from Λ onto $\Omega_t := \boldsymbol{\varphi}_t(\Lambda)$, that we call (*deformed*) *body at time t*. For each t, we then define also the inverse diffeomorphism $\boldsymbol{\varphi}_t^{-1} : \Omega_t \to \Lambda$. The variables $\boldsymbol{\lambda} \in \Lambda$ are called *material* or *Lagrangian coordinates*, whereas $\boldsymbol{x}$ are usually referred to as *spatial* or *Eulerian coordinates*.

The local deformation of the body is measured by the *deformation gradient*

$$\mathsf{F}(\boldsymbol{\lambda}, t) = \mathsf{F}_t(\boldsymbol{\lambda}) := \nabla \boldsymbol{\varphi}(\boldsymbol{\lambda}, t),$$

that maps tangent vectors to the material manifold into tangent vectors to the body Ω_t. Its determinant

$$J_t := \det \mathsf{F}_t$$

quantifies the local change of volume induced by the motion. It is customary to assume that the deformation gradient maps a basis of the tangent space to Λ into a basis of the tangent space to Ω_t with the same orientation. This translates into the orientation-preserving condition $J(\boldsymbol{\lambda}, t) > 0$ for every $\boldsymbol{\lambda} \in \Lambda$ and for every time t. Moreover, assuming that material points and their mass are neither created nor destroyed, we have mass conservation, that implies the relation

$$\hat{\rho}(\boldsymbol{\lambda}) = \rho(\boldsymbol{\varphi}_t(\boldsymbol{\lambda}), t)\, J(\boldsymbol{\lambda}, t), \tag{3}$$

where $\hat{\rho}(\boldsymbol{\lambda})$ and $\rho(\boldsymbol{x}, t)$ are the mass density on Λ and Ω_t, respectively.

A second fundamental quantity in the kinematic description of a continuum is the velocity field, which can appear in two versions: the *material velocity field* $\frac{\partial \boldsymbol{\varphi}}{\partial t}(\boldsymbol{\lambda}, t)$ and the *spatial velocity field* $\boldsymbol{v}(\boldsymbol{x}, t)$. The two are connected by the nonlinear ordinary differential equation

$$\frac{\partial \boldsymbol{\varphi}}{\partial t}(\boldsymbol{\lambda}, t) = \boldsymbol{v}(\boldsymbol{\varphi}(\boldsymbol{\lambda}, t), t). \tag{4}$$

We also introduce the so-called *material derivative* of a generic spatial vector field $\boldsymbol{f}$

$$\mathcal{D}_{\boldsymbol{v}} \boldsymbol{f} = \frac{\partial \boldsymbol{f}}{\partial t} + (\boldsymbol{v} \cdot \nabla)\boldsymbol{f},$$

that simply arises from calculating the time derivative of $\boldsymbol{f}(\boldsymbol{\varphi}_t(\boldsymbol{\lambda}), t)$ exploiting the chain rule. In particular, while the material acceleration is simply the second time derivative of $\boldsymbol{\varphi}_t$, the spatial acceleration field is defined by

$$\boldsymbol{a} := \mathcal{D}_{\boldsymbol{v}} \boldsymbol{v} = \frac{\partial \boldsymbol{v}}{\partial t} + (\boldsymbol{v} \cdot \nabla)\boldsymbol{v}.$$

This expression represents one of the central nonlinear features of continuum mechanics.

4 Material version of the linear momentum balance

In order to express the principle of virtual works in the continuum setting, we have to identify a suitable set of virtual velocities. In analogy to the discrete case, the virtual velocity will be a vector belonging to the tangent space $T_{\boldsymbol{\varphi}_t(\boldsymbol{\lambda})}\Omega_t$, for some label $\boldsymbol{\lambda}$ and time t. Thus, we consider virtual velocities as given by the composition of a spatial field $\boldsymbol{w}(\boldsymbol{x},t)$ with the placement, i.e. material fields of the form $\boldsymbol{w}(\boldsymbol{\varphi}_t(\boldsymbol{\lambda}),t)$. The principle of virtual works states that any real motion $\boldsymbol{\varphi}_t(\boldsymbol{\lambda})$ on a time interval $[t_i,t_f]$ of a continuum with material manifold Λ and mass density $\hat{\rho}$ is such that

$$\int_{t_i}^{t_f}\int_{\Lambda}\Big(-\hat{\rho}(\boldsymbol{\lambda})\frac{\partial\boldsymbol{\varphi}}{\partial t}(\boldsymbol{\lambda},t)\cdot\frac{d}{dt}\boldsymbol{w}(\boldsymbol{\varphi}_t(\boldsymbol{\lambda}),t)+ \\ -\hat{\rho}(\boldsymbol{\lambda})\hat{\boldsymbol{b}}\cdot\boldsymbol{w}(\boldsymbol{\varphi}_t(\boldsymbol{\lambda}),t)+\mathsf{P}:\frac{\partial}{\partial\boldsymbol{\lambda}}\boldsymbol{w}(\boldsymbol{\varphi}_t(\boldsymbol{\lambda}),t)\Big)d\boldsymbol{\lambda}dt=0, \tag{5}$$

for any admissible virtual velocity field $\boldsymbol{w}(\boldsymbol{\varphi}_t(\boldsymbol{\lambda}),t)$.

In equation (5), we find the material linear momentum density $\hat{\rho}\frac{\partial\boldsymbol{\varphi}}{\partial t}$, the material force density per unit mass $\hat{\boldsymbol{b}}$ and the novel object P, that is the *first Piola-Kirchhoff stress tensor* and represents the internal stresses within the body. A continuum model, to be completely identified, will require also the specification of the dependence of

$$\hat{\boldsymbol{b}}\left(\boldsymbol{\lambda},t,\boldsymbol{\varphi}_t(\boldsymbol{\lambda}),\frac{\partial\boldsymbol{\varphi}}{\partial t}(\boldsymbol{\lambda},t),\frac{\partial\boldsymbol{\varphi}}{\partial\boldsymbol{\lambda}}(\boldsymbol{\lambda},t),\dots\right)$$

and

$$\mathsf{P}\left(\boldsymbol{\lambda},t,\boldsymbol{\varphi}_t(\boldsymbol{\lambda}),\frac{\partial\boldsymbol{\varphi}}{\partial t}(\boldsymbol{\lambda},t),\frac{\partial\boldsymbol{\varphi}}{\partial\boldsymbol{\lambda}}(\boldsymbol{\lambda},t),\dots\right)$$

on the kinematic quantities. Note that : in (5) denotes the matrix product $\mathsf{A}:\mathsf{C}=\mathrm{tr}(\mathsf{A}\mathsf{C}^\top)$.

We now want to derive a local version of the linear momentum balance. In order to do so, we start from the principle of virtual work (5), which represents a weak form of the balance equation, integrate by parts in time the first term and apply the divergence theorem to the last term, to obtain

$$\int_{t_i}^{t_f}\int_{\Lambda}\left(\hat{\rho}\frac{\partial^2\boldsymbol{\varphi}}{\partial t^2}-\hat{\rho}\hat{\boldsymbol{b}}-\mathrm{div}\,\mathsf{P}\right)\cdot\boldsymbol{w}(\boldsymbol{\varphi}_t(\boldsymbol{\lambda}),t)\,d\boldsymbol{\lambda}dt+ \\ +\int_{t_i}^{t_f}\int_{\partial\Lambda}\mathsf{P}\boldsymbol{m}\cdot\boldsymbol{w}(\boldsymbol{\varphi}_t(\boldsymbol{\lambda}),t)\,dSdt=0, \tag{6}$$

where $\boldsymbol{m}$ denotes the unit outer normal to the boundary $\partial\Lambda$, and dS the appropriate Hausdorff measure. Since (6) holds for any smooth virtual velocity with vanishing initial and final conditions (for simplicity), and since the integrands are continuous, thanks to the fundamental lemma of calculus of variations we can conclude that

$$\hat{\rho}\frac{\partial^2\boldsymbol{\varphi}}{\partial t^2}=\mathrm{div}\,\mathsf{P}+\hat{\rho}\hat{\boldsymbol{b}},$$

which is the *local balance of linear momentum in material coordinates*. The vanishing of the surface integral is related to boundary conditions. We can partition $\partial\Lambda$ as the disjoint

union of S_D and S_N. On S_D we can impose the so-called Dirichlet boundary conditions, according to which the placement value is fixed and therefore virtual velocities vanish on S_D. On S_N we can impose Neumann boundary conditions letting the virtual velocity be non-zero, and obtaining then the local traction-free condition

$$\mathsf{P}(\boldsymbol{\lambda},t)\boldsymbol{m}(\boldsymbol{\lambda}) = \mathbf{0} \qquad \text{for } \boldsymbol{\lambda} \in S_N.$$

5 Spatial version of the linear momentum balance

Since it is very often easier to think about forces and stresses in the ambient space, we now derive also the spatial version of the principle of virtual works. By performing the change of variables $\boldsymbol{\lambda} = \boldsymbol{\varphi}_t^{-1}(\boldsymbol{x})$ in (5), and recalling equations (3) and (4), one gets

$$\int_{t_i}^{t_f}\int_{\Omega_t}\Big(-\rho(\boldsymbol{x},t)\boldsymbol{v}(\boldsymbol{x},t)\cdot\mathcal{D}_{\boldsymbol{v}}\boldsymbol{w}(\boldsymbol{x},t) - \rho(\boldsymbol{x},t)\boldsymbol{b}(\boldsymbol{x},t)\cdot\boldsymbol{w}(\boldsymbol{x},t) + \\ +\mathsf{T}(\boldsymbol{x},t) : \nabla\boldsymbol{w}(\boldsymbol{x},t)\Big)d\boldsymbol{x}dt = 0, \tag{7}$$

where we introduced the spatial force density

$$\boldsymbol{b}(\boldsymbol{x},t) = \hat{\boldsymbol{b}}\left(\boldsymbol{\lambda},t,\boldsymbol{\varphi}_t(\boldsymbol{\lambda}),\frac{\partial\boldsymbol{\varphi}}{\partial t}(\boldsymbol{\lambda},t),\frac{\partial\boldsymbol{\varphi}}{\partial\boldsymbol{\lambda}}(\boldsymbol{\lambda},t),\dots\right)\Big|_{\boldsymbol{\lambda}=\boldsymbol{\varphi}_t^{-1}(\boldsymbol{x})}$$

and the *Cauchy stress tensor*

$$\mathsf{T}(\boldsymbol{x},t) = \frac{1}{J_t(\boldsymbol{\lambda})}\mathsf{P}\left(\boldsymbol{\lambda},t,\boldsymbol{\varphi}_t(\boldsymbol{\lambda}),\frac{\partial\boldsymbol{\varphi}}{\partial t}(\boldsymbol{\lambda},t),\frac{\partial\boldsymbol{\varphi}}{\partial\boldsymbol{\lambda}}(\boldsymbol{\lambda},t),\dots\right)\mathsf{F}_t^{\top}(\boldsymbol{\lambda})\Big|_{\boldsymbol{\lambda}=\boldsymbol{\varphi}_t^{-1}(\boldsymbol{x})}.$$

The principle of virtual works in spatial formulation states that (7) holds true for every admissible virtual velocity $\boldsymbol{w}(\boldsymbol{x},t)$. However, we should emphasize that this formulation has a more complex structure, because the domain Ω_t is itself part of the solution, since it evolves in time. For this reason, the spatial formulation is preferred only in cases where the domain does not depend on time. We shall see in the last section that there are relevant situations in which this is the case.

Finally, we can start from (7) and repeat the procedure introduced in the previous section to find the local balance of linear momentum. Performing an integration by parts, recalling the definition of material derivative, and applying the divergence theorem, we obtain

$$\int_{t_i}^{t_f}\int_{\Omega_t}(\rho\mathcal{D}_{\boldsymbol{v}}\boldsymbol{v} - \rho\boldsymbol{b} - \operatorname{div}\mathsf{T})\cdot\boldsymbol{w}\,d\boldsymbol{x}dt + \int_{t_i}^{t_f}\int_{\partial\Omega_t}\mathsf{T}\boldsymbol{n}\cdot\boldsymbol{w}dSdt + \\ -\int_{t_i}^{t_f}\int_{\partial\Omega_t}\rho(\boldsymbol{v}\cdot\boldsymbol{n})\boldsymbol{v}\cdot\boldsymbol{w}dSdt = 0, \tag{8}$$

where $\boldsymbol{n}$ is the outer unit normal to $\partial\Omega_t$ and where we also exploited the so-called *continuity equation* (that is the local spatial expression of mass conservation)

$$\mathcal{D}_{\boldsymbol{v}}\rho(\boldsymbol{x},t) + \rho(\boldsymbol{x},t)\operatorname{div}\boldsymbol{v}(\boldsymbol{x},t) = 0, \tag{9}$$

Since (8) holds for any smooth virtual velocity with vanishing initial and final conditions (for simplicity), and since the integrands are continuous, thanks to the fundamental lemma of calculus of variations we can conclude that

$$\rho \mathcal{D}_{\boldsymbol{v}} \boldsymbol{v} = \operatorname{div} \mathsf{T} + \rho \boldsymbol{b} \tag{10}$$

which is the *local balance of linear momentum in spatial coordinates*. As for the boundary conditions, if we partition the boundary $\partial\Omega_t$ in a Dirichlet part Γ_D and a Neumann part Γ_N, we can impose $\boldsymbol{v} = 0$ on Γ_D and $\mathsf{T}\boldsymbol{n} = \boldsymbol{0}$ on Γ_N. The last boundary integral in (8) highlights the fact that the spatial formulation is better suited to treat situations in which the region of space occupied by the continuum does not depend on time. In fact, that term vanishes when either $\boldsymbol{v} = \boldsymbol{0}$ or $\boldsymbol{v} \cdot \boldsymbol{n} = 0$, both cases are consistent with a time-independent Ω_t.

6 Physical interpretation of the stress tensor

To have a better understanding of the role of the Cauchy stress, we consider an open subset $\hat{\Lambda} \subset \Lambda$, with smooth boundary, and its image $\hat{\Omega}_t := \boldsymbol{\varphi}_t(\hat{\Lambda})$. We want to compute the time derivative of the total linear momentum of points in $\hat{\Omega}_t$, exploiting the following

Theorem 6.1 (Reynolds' transport theorem) *For any sufficiently regular spatial field $F(\boldsymbol{x}, t)$ and smooth $\hat{\Omega}_t := \boldsymbol{\varphi}_t(\hat{\Lambda})$, one has*

$$\frac{d}{dt} \int_{\hat{\Omega}_t} F(\boldsymbol{x}, t) d\boldsymbol{x} = \int_{\hat{\Omega}_t} (\mathcal{D}_{\boldsymbol{v}} F(\boldsymbol{x}, t) + F(\boldsymbol{x}, t) \operatorname{div} \boldsymbol{v}(\boldsymbol{x}, t)) d\boldsymbol{x}.$$

This result, combined with the continuity equation (9) and the momentum balance (10), yields

$$\frac{d}{dt} \int_{\hat{\Omega}_t} \rho \boldsymbol{v} \, d\boldsymbol{x} = \int_{\hat{\Omega}_t} \rho \boldsymbol{b} \, d\boldsymbol{x} + \int_{\partial\hat{\Omega}_t} \mathsf{T}\boldsymbol{n} \, dS.$$

We thus see that the time variation of the total linear momentum of the region $\hat{\Omega}_t$ equals the total body force (volume integral) acting on points in $\hat{\Omega}_t$ plus a surface integral involving the Cauchy stress and the normal $\boldsymbol{n}$ to $\partial\hat{\Omega}_t$. Therefore, we can interpret $\mathsf{T}\boldsymbol{n}$ as encoding the momentum flux per unit time through the boundary of any smooth subdomain. Moreover, we can describe the vector $\mathsf{T}\boldsymbol{n}$ as a contact force acting on the same boundary. The stress tensor is then a linear operator that, for any direction normal to a surface drawn within the body, gives the force acting locally on that surface due to contact interactions. Internal forces in the form of contact interactions are a peculiarity of continuum theories and their description is what characterizes different classes of materials. As an example, we can consider the sea as the body, and the subdomain $\hat{\Omega}_t$ to be a region of water, the external volume force is then the gravitational force due to the mass of the region, while the internal (contact) force is the pressure experienced by the region due to the surrounding water.

7 Infinite-dimensional setting for the motion of an incompressible perfect fluid

As discussed in the previous sections, the basic descriptor of the continuum is the placement

$$\boldsymbol{\varphi}_t : \Lambda \to \mathbb{R}^3.$$

Thus, the evolution of a continuum can be viewed as a curve in the space of deformations, assigning to each instant of time a deformation map. These deformation maps belong to a function space, which is inherently infinite-dimensional. Therefore, a natural question is whether studying the geometric properties of these spaces of admissible deformations can provide information about the motions of a continuum. In their seminal works, first Arnold [1] and then Ebin and Marsden [3] developed a geometric description of the motion of incompressible perfect fluids based on groups of diffeomorphisms.

In order to fully understand their geometric approach, we first introduce the basic notions regarding incompressible perfect fluids. A fluid is said to be perfect if it is homogeneous, i.e. its material density is constant, and inviscid, i.e. it has no viscosity. We further restrict our attention to incompressible fluids, for which only volume-preserving motions are admissible, i.e. the volume of the region occupied by the fluid does not change over time. This translates to the condition

$$J_t = 1 \qquad \text{for every } t.$$

or, in spatial coordinates,

$$\operatorname{div} \boldsymbol{v}_t = 0 \qquad \text{for every } t.$$

In the case of an incompressible perfect fluid, it is reasonable to think that the fluid moves in a fixed region of space Ω (namely, the channel within which the fluid is flowing). Consequently, the placement map can be regarded as a volume-preserving diffeomorphism

$$\boldsymbol{\varphi}_t : \Omega \to \Omega.$$

In this setting, the spatial formulation introduced in the previous sections is particularly convenient, since the domain Ω is fixed in time. Homogeneity and incompressibility imply that the mass density remains constant both in space and in time. In the following, we will therefore assume it to be equal to one, for the sake of simplicity.

Classically, to specify a continuum model one needs to prescribe a constitutive law for the stress tensor, which encodes the characteristics of the material. In the case of incompressible perfect fluids, this constitutive equation is

$$\mathsf{T}(\boldsymbol{x}, t) = -p(\boldsymbol{x}, t)\,\mathsf{I}$$

where p is the scalar pressure field and I is the identity matrix. Plugging this constitutive equation into (10) yields the equation of motion for an incompressible perfect fluid

$$\partial_t \boldsymbol{v} + (\boldsymbol{v} \cdot \nabla)\boldsymbol{v} = -\nabla p, \tag{11}$$

that is the celebrated *Euler's equation*. The geometric approach introduced in [1], and further developed in [3], is based on the study of the properties of the space

$$\mathcal{D}^s := \{\boldsymbol{\varphi} \in H^s(\Omega, \Omega) \mid \boldsymbol{\varphi} \text{ is bijective and } \boldsymbol{\varphi}^{-1} \in H^s(\Omega, \Omega)\}$$

with $s \geq 3$. The choice of Sobolev regularity is motivated by the Hilbert structure of $H^s(\Omega, \Omega)$, which provides a convenient analytical framework. In particular, maps in $H^s(\Omega, \Omega)$ have square-integrable derivatives up to order s. Moreover, choosing $s \geq 3$ ensures, by Sobolev's embedding theorem, that the H^s-topology is stronger than the C^1-topology. These properties are crucial to show that $\mathcal{D}^s$ is a differentiable manifold and a topological group. Since we are interested in incompressible perfect fluids, we restrict our attention to the subset

$$\mathcal{D}^s_\mu = \{\boldsymbol{\varphi} \in \mathcal{D}^s \mid \det \nabla \boldsymbol{\varphi} = 1\}$$

of volume-preserving diffeomorphisms. It is possible to prove the following fundamental

Theorem 7.1 *Let $\Omega \subset \mathbb{R}^3$ be a bounded connected domain with Lipschitz boundary and let $s \geq 3$ be an integer. The Jacobian determinant map*

$$\begin{aligned} J : \mathcal{D}^s &\to H^{s-1}(\Omega, \mathbb{R}) \\ \boldsymbol{\varphi} &\mapsto \det \nabla \boldsymbol{\varphi} \end{aligned}$$

is a submersion on $\mathcal{D}^s_\mu$. In particular, the set of volume-preserving deformations $\mathcal{D}^s_\mu$ is a submanifold of $\mathcal{D}^s$, with

$$T_{\boldsymbol{\varphi}} \mathcal{D}^s_\mu = \{\boldsymbol{v} \circ \boldsymbol{\varphi} \in H^s(\Omega, \mathbb{R}^3) \mid \operatorname{div} \boldsymbol{v} = 0\}.$$

The key insight of the Arnold-Ebin-Marsden framework is that Euler's equation (11) can be interpreted as the geodesic equation on the submanifold $\mathcal{D}^s_\mu \subset \mathcal{D}^s$ of volume-preserving deformations, endowed with the (weak) Riemannian structure induced by the metric

$$\langle \boldsymbol{U}, \boldsymbol{W} \rangle_{\boldsymbol{\varphi}} = \int_\Omega \boldsymbol{U}(\boldsymbol{\lambda}) \cdot \boldsymbol{W}(\boldsymbol{\lambda}) \, d\boldsymbol{\lambda},$$

where $\boldsymbol{U} := \boldsymbol{u} \circ \boldsymbol{\varphi}, \boldsymbol{W} := \boldsymbol{w} \circ \boldsymbol{\varphi} \in T_{\boldsymbol{\varphi}} \mathcal{D}^s_\mu$. In fact, a geodesic on the submanifold $\mathcal{D}^s_\mu \subset \mathcal{D}^s$ with respect to this metric can be characterized as a curve $\boldsymbol{\varphi} : [0, T] \to \mathcal{D}^s_\mu$ that minimizes the functional

$$\mathcal{F}[\boldsymbol{\varphi}] := \frac{1}{2} \int_0^T \langle \dot{\boldsymbol{\varphi}}(t), \dot{\boldsymbol{\varphi}}(t) \rangle_{\boldsymbol{\varphi}(t)} dt = \int_0^T \int_\Omega \frac{1}{2} |\dot{\boldsymbol{\varphi}}(\boldsymbol{\lambda}, t)|^2 d\boldsymbol{\lambda}, \tag{12}$$

that corresponds to the kinetic energy of the fluid. In order to find stationary curves of the functional (12), we need to compute

$$\delta \mathcal{F} = \frac{d}{ds} \mathcal{F}[\boldsymbol{\varphi} + s\boldsymbol{W}]|_{s=0} = 0, \tag{13}$$

with $\boldsymbol{W} := \boldsymbol{w} \circ \boldsymbol{\varphi} \in T_{\boldsymbol{\varphi}}\mathcal{D}^s_\mu$. Equation (13) becomes

$$
\begin{aligned}
0 &= \frac{d}{ds}\int_0^T \int_\Omega \frac{1}{2}\left|\frac{\partial}{\partial t}\left(\boldsymbol{\varphi}(\boldsymbol{\lambda},t) + s\boldsymbol{w}(\boldsymbol{\varphi}(\boldsymbol{\lambda}),t)\right)\right|^2 d\boldsymbol{\lambda}dt|_{s=0} \\
&= \int_0^T \int_\Omega \left(\boldsymbol{v}(\boldsymbol{x},t)\cdot \mathcal{D}_{\boldsymbol{v}}\boldsymbol{w}(\boldsymbol{x},t)\right) d\boldsymbol{x}dt = \int_0^T \int_\Omega -\mathcal{D}_{\boldsymbol{v}}\boldsymbol{v}(\boldsymbol{x},t)\cdot \boldsymbol{w}(\boldsymbol{x},t) d\boldsymbol{x}dt,
\end{aligned} \tag{14}
$$

for every $\boldsymbol{w}$ such that $\operatorname{div}\boldsymbol{w} = 0$. It is well known that, for every $t \in \mathbb{R}$, the gradient of a scalar field $p(\cdot,t)$ is orthogonal (in the L^2-sense) to any divergence-free vector field $\boldsymbol{w}(\cdot,t)$ that vanishes on the boundary of Ω. In fact, let

$$
\boldsymbol{h}(\boldsymbol{x},t) = \nabla_{\boldsymbol{x}} p(\boldsymbol{x},t),
$$

from the divergence theorem and the properties of the fields $\boldsymbol{w}(\cdot,t)$ and $p(\cdot,t)$, it follows that

$$
\begin{aligned}
&\int_\Omega \boldsymbol{w}(\boldsymbol{x},t)\cdot\boldsymbol{h}(\boldsymbol{x},t)d\boldsymbol{x} = \int_\Omega \boldsymbol{w}(\boldsymbol{x},t)\cdot\nabla_{\boldsymbol{x}} p(\boldsymbol{x},t)d\boldsymbol{x} \\
&= -\int_\Omega p(\boldsymbol{x},t)\operatorname{div}_{\boldsymbol{x}}\boldsymbol{w}(\boldsymbol{x},t)d\boldsymbol{x} + \int_{\partial\Omega} p(\boldsymbol{x},t)\boldsymbol{n}\cdot\boldsymbol{w}(\boldsymbol{x},t)\, dS = 0,
\end{aligned}
$$

where $\boldsymbol{n}$ is the outward unit normal to $\partial\Omega$ and dS denotes the integration with respect to the two-dimensional Hausdorff measure on $\partial\Omega$. Since the L^2-orthogonal complement of divergence-free vector fields consists of gradients, equation (14) implies that

$$
\mathcal{D}_{\boldsymbol{v}}\boldsymbol{v} = -\nabla p,
$$

that is exactly Euler's equation (11). Therefore, the motion of an incompressible perfect fluid can be identified with the geodesic flow on the infinite-dimensional Riemannian manifold $\mathcal{D}^s_\mu$ of volume-preserving diffeomorphisms.

8 Conclusion

The transition from Newtonian mechanics to continuum mechanics naturally leads from finite-dimensional configuration spaces to spaces of functions, which are inherently infinite-dimensional. Through the principle of virtual works, the balance laws governing continua can be formulated in a way that closely parallels the finite-dimensional case while remaining applicable to systems with infinitely many degrees of freedom. Within this framework, the placement map emerges as the fundamental kinematic variable, and the motion can be viewed as a curve in the collection of admissible deformations. An example of how the geometry of these infinite-dimensional spaces can be related to continuum motions is the seminal Arnold-Ebin-Marsden interpretation of Euler's equation of motion for an incompressible perfect fluid as the geodesic equation on the manifold of volume-preserving diffeomorphisms. This structure allows one to recast the analytical problem of fluid motion into a geometric problem of finding geodesics, to which the methods of global analysis can be applied. This geometric formulation has since become a cornerstone in the modern

understanding of hydrodynamics, as it connects the analytical structure of the equations of motion with the differential geometry of the underlying configuration space. More broadly, it illustrates how geometric structures can provide insight into the qualitative behaviour of physical systems and motivates the study of infinite-dimensional manifolds arising in continuum theories.

References

[1] V. Arnold, *Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits.* Ann. Inst. Fourier (Grenoble), 16: 319–361, 1966.

[2] A.J. Chorin and J.E. Marsden, "A mathematical introduction to fluid mechanics". Volume 4 of Texts in Applied Mathematics. Springer-Verlag, New York, third edition, 1993.

[3] D.G. Ebin and J.E. Marsden, *Groups of diffeomorphisms and the motion of an incompressible fluid.* Ann. of Math. (2), 92: 102–163, 1970.

[4] M.E. Gurtin, "An introduction to continuum mechanics". Volume 158 of Mathematics in Science and Engineering. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London, 1981.

[5] J.E. Marsden and T.J.R. Hughes, "Mathematical foundations of elasticity". Dover Publications, Inc., New York, 1994. Corrected reprint of the 1983 original.

Isoperimetric Inequalities Between Two Fractional Perimeters

Giacomo Cozzi (*)

Abstract. In this note we will introduce the notion of fractional perimeter of a set, show its main properties and motivate the interest in this object. Next, we will present the results known in the literature concerning isoperimetric inequalities with fractional perimeters. Finally, we will discuss a project in which we prove that the ball is a local minimizer of the scale invariant isoperimetric ratio between two fractional perimeters under C^1 small graph perturbations: this parallels the results known in the literature where one of the two fractional perimeters is replaced by the volume.
This is a joint work with Annalisa Massaccesi (University of Padova), Giovanni Alberti and Jeremy Mirmina (University of Pisa).

1 Fractional perimeter

We begin this section with the definition of fractional perimeter

Definition 1 Let $\alpha \in (0,1)$, and $E \subset \mathbb{R}^n$. Then we set

$$P_\alpha(E) := \iint_{E\times E^c} \frac{dx\,dy}{|x-y|^{n+\alpha}} \tag{1.1}$$

to be the α-perimeter of E.

The above definition appeared for the first time in [4], where it was also connected with the notion of nonlocal minimal surface i.e., of a subset of $\mathbb{R}^n$ which minimizes a properly localized version of (1.1) in any compact set $\Omega \subset \mathbb{R}^n$. Nowadays nonlocal minimal surfaces are a central object in geometric analysis due to their role as approximations of minimal surfaces.

Roughly speaking the quantity in (1.1) is a superposition of all the interaction occurring between points inside E and outside E (namely, in E^c). The contribution of each couple of points to $P_\alpha(E)$ depends on the distance between the two points: the smaller the distance, the bigger the contribution. Observing this, one can already guess why to (1.1) it was given

(*)Ph.D. course, Università di Padova, Dip. Matematica, via Trieste 63, I-35121 Padova, Italy. E-mail: cozzi@math.unipd.it . Seminar held on 9 April 2026.

the name " (fractional) perimeter": the most relevant contributions are given by pairs of points living near the boundary ∂E (see below for further details).

Remark 1 We remark here some properties of the α-perimeter

i) A priori, the positive quantity (1.1) could be $+\infty$. In the following subsection we will treat with more detail in which cases this can be excluded.

ii) For any $\lambda > 0$, it holds

$$P_\alpha(\lambda E) = \lambda^{n-\alpha} P_\alpha(E). \tag{1.2}$$

This can be checked simply applying the change of variables $x' = \lambda x$, $y' = \lambda y$ in (1.1). It is interesting, though, to observe that the homogeneity of the α-perimeter ranges between the one of the euclidean volume $|\cdot|$ (for which $|\lambda E| = \lambda^n |E|$) and the one of the perimeter (for which $P(\lambda E) = \lambda^{n-1} P(E)$). Again, we refer to the next subsection for the precise connection between these objects.

iii) If E and F are disjoint subsets with finite α-perimeter, then

$$P_\alpha(E \cup F) = P_\alpha(E) + P_\alpha(F) - 2 \iint_{E\times F} \frac{dxdy}{|x-y|^{n+\alpha}} \leq P_\alpha(E) + P_\alpha(F), \tag{1.3}$$

namely, the fractional perimeter is sub-additive with respect to disjoint union.

iv) A set E is a set of finite perimeter (in the De Giorgi - Caccioppoli sense if the function χ_E is in the space of functions with bounded variation $BV(\mathbb{R}^n)$, or equivalently, if its total variation $|D\chi_E|$ is a finite Radon measure over $\mathbb{R}^n$ (see for example [13, Chapter 12] for further details). If this is the case, we can show that

$$P_\alpha(E) < +\infty \quad \text{for any } \alpha \in (0,1).$$

Indeed, for a certain function $u \in W^{1,1}(\mathbb{R}^n) \subset BV$, it holds

$$\begin{aligned}
&\iint \frac{|u(x)-u(y)|}{|x-y|^{n+\alpha}} = \iint \frac{|u(y+z)-u(y)|}{|z|^{n+\alpha}}\, dy\, dz \\
&= \int\int_{B_1(0)} \frac{|u(y+z)-u(y)|}{|z|^{n+\alpha}}\, dy\, dz + \int\int_{B_1(0)^c} \frac{|u(y+z)-u(y)|}{|z|^{n+\alpha}}\, dy\, dz \\
&\leq \int\int_{B_1(0)} \int_0^1 \frac{|\nabla u(y+tz)\cdot z|}{|z|^{n+\alpha}} dt\, dz\, dy + 2\|u\|_{L^1(\mathbb{R}^n)} \int_{B_1^c(0)} \frac{dz}{|z|^{n+\alpha}} \\
&\leq \int_{B_1(0)} \frac{1}{|z|^{n+\alpha-1}} \int_0^1 \left(\int |\nabla u(y+tz)|\, dy \right) dt\, dz + 2\|u\|_{L^1(\mathbb{R}^n)} \int_{B_1^c(0)} \frac{dz}{|z|^{n+\alpha}} \\
&\leq \|\nabla u\|_{L^1(\mathbb{R}^n)} \int_{B_1(0)} \frac{dz}{|z|^{n+\alpha-1}} + 2\|u\|_{L^1(\mathbb{R}^n)} \int_{B_1^c(0)} \frac{dz}{|z|^{n+\alpha}} \\
&\leq c_{n,\alpha} \|u\|_{W^{1,1}(\mathbb{R}^n)}.
\end{aligned}$$

Moreover, for a function $u \in BV(\mathbb{R}^n)$, consider an approximation $(u_k)_{k\in\mathbb{N}} \subset BV(\mathbb{R}^n) \cap C^\infty(\mathbb{R}^n)$ such that

$$u_k \xrightarrow[L^1]{} u \quad \text{and} \quad \lim_{k\to\infty} \int_{\mathbb{R}^n} |\nabla u_k| \, dx = |Du|(\mathbb{R}^n).$$

Then, by Fatou Lemma, we have

$$\begin{aligned} \|u\|_{BV(\mathbb{R}^n)} &= \lim_{k\to\infty} \|u_k\|_{W^{1,1}(\mathbb{R}^n)} \\ &= \frac{1}{c_{n,\alpha}} \lim_{k\to\infty} \iint \frac{|u_k(x) - u_k(y)|}{|x-y|^{n+\alpha}} \, dx \, dy \\ &\geq \frac{1}{c_{n,\alpha}} \iint \frac{|u(x) - u(y)|}{|x-y|^{n+\alpha}} \, dx \, dy. \end{aligned}$$

Applying this to $u = \chi_E$ and noticing that (1.1) could be rewritten as

$$P_\alpha(E) = \iint_{E\times E^c} \frac{dx\, dy}{|x-y|^{n+\alpha}} = \frac{1}{2} \iint_{\mathbb{R}^n\times\mathbb{R}^n} \frac{|\chi_E(x) - \chi_E(y)|}{|x-y|^{n+\alpha}} \, dx \, dy,$$

we obtain the claim.

1.1 Perimeter, α-perimeter and volume

The following two theorems clarify the connections between the notions of Perimeter and α-perimeter, and between α-perimeter and volume, respectively.

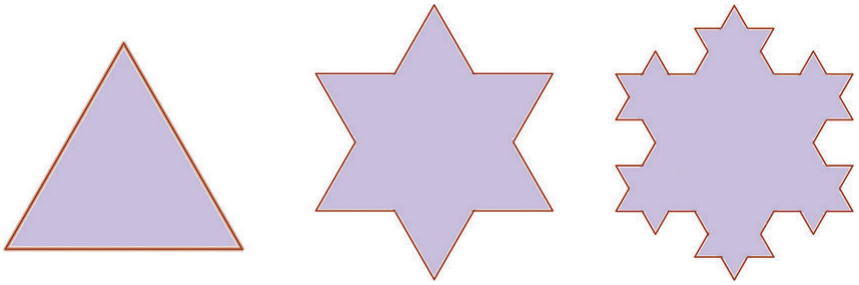
Theorem 1 ([2], [7]) *Let E be bounded. Then*

$$P(E) < +\infty \iff P_\alpha(E) < +\infty \quad \textit{for any } \alpha \in (0,1) \textit{ and } \liminf_{\alpha\to 1} \frac{(1-\alpha)}{\omega_n} P_\alpha(E) < +\infty.$$

Theorem 2 ([14]) *Let E be bounded. Then*

$$\lim_{\alpha\to 0} \frac{\alpha}{\omega_n} P_\alpha(E) = |E|.$$

In what follows, we will present an example that shows that not all sets with finite fractional perimeter must have finite perimeter: consider E to be the Von Koch's snowflake, and define $(E_k)_{k\in\mathbb{N}}$ the sequence of which it is a limit, as shown here below.



In order to pass from E_k to E_{k+1}, we need to replace any edge of the set E_k with a curve made of four segments each of which measures $1/3$ of the original edge. Therefore we deduce that

$$P(E_{k+1}) = \frac{4}{3}P(E_k),$$

and hence

$$P(E) = \lim_{k\to\infty} P(E_k) = +\infty.$$

On the other hand, we have that $E_1 = E_0 \cup (E_1^1 \cup E_1^2 \cup E_1^3)$, where E_1^i are the three small triangles with edge of size $1/3$ of the original edge of E_0. When passing from E_1 to E_2, instead, we need to add twelve small triangles with edge of size $1/9$ of the original one. In order to pass from E_k to E_{k+1} we need to add 3×4^k triangles with edge of size $3^{-(k+1)}$ of the original edge. Since all these union of triangles are disjoint, by (1.2), we deduce that, for any $\alpha \in (0,1)$

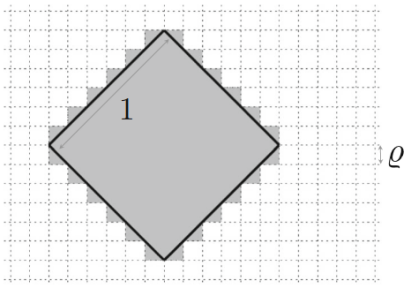
$$\begin{aligned} P_\alpha(E) &\le P_\alpha(E_0) + 3\sum_{k=0}^{\infty} 4^k P_\alpha\left(\frac{E_0}{3^{k+1}}\right) \\ &= P_\alpha(E_0) + 3P_\alpha(E_0)\sum_{k=0}^{\infty}\frac{4^k}{3^{(k+1)(2-\alpha)}} \\ &= P_\alpha(E_0) + 3^{3-\alpha}P_\alpha(E_0)\sum_{k=0}^{\infty}\left(\frac{4}{3^{\alpha-1}}\right)^k. \end{aligned}$$

We deduce that $P_\alpha(E) < +\infty$ as soon as $\alpha < 2 - \ln 4/\ln 3$.

Remark 2 We could also construct more sophisticated examples showing that there exist sets E with finite α-perimeter for any $\alpha \in (0,1)$, yet such that $P(E) = +\infty$: a classic one is the set $E = \bigcup_{k\in\mathbb{N}} I_{2k} \subset \mathbb{R}$, where $I_k = (b^{k+1}, b^k)$ for a fixed $b \in (0,1)$.

1.2 Motivations for fractional perimeters

To conclude the first section, we illustrate an example contained in [6] that shows how, in some situations, fractional perimeters can be used to construct models which are more robust and well behaved than those obtained using the classical perimeter. Fix $\rho > 0$ and suppose you want to approximate a square Q rotated by 45 degrees just using the pixels of the grid, as shown in the picture below.



Independently on how small ρ (i.e., on how good the resolution of your screen is), the classical perimeter is an imprecise tool to measure how good your approximation actually is. Indeed, if we call A the grey area in the figure above, we have that $P(A) = \sqrt{2}P(Q)$. On the other hand, for any $\alpha \in (0,1)$ it holds

$$\begin{aligned} |P_\alpha(A) - P_\alpha(Q)| &\leq \iint_{(A\setminus Q)\times(A\setminus Q)^c} \frac{dx\,dy}{|x-y|^{n+\alpha}} = P_\alpha(A \setminus Q) \\ &\leq \sum_i P_\alpha(T_i), \end{aligned} \tag{1.4}$$

where T_i denotes a generic triangle lying on one edge of Q, and we applied (1.3). Since the size of T_i is of the order ρ and the number of triangles on the edge of Q is of the order ρ^{-1}, applying Equation (1.2) in (1.4), it follows

$$|P_\alpha(A) - P_\alpha(Q)| \lesssim \rho^{-1}\rho^{2-\alpha} \xrightarrow{\rho\to 0} 0.$$

The "nonlocality" of the fractional perimeter, opposed to the sensitivity of the perimeter to small discrepancies, leads to a way to properly measure how good the approximation A was in terms of a quantity that is not the perimeter, but still can approximate it as $\alpha \to 1$ (cf. Theorem 1).

Apart from this example, many other motivations for studying fractional perimeters are nowadays widely known. Among them, the understanding of steady states for nonlinear interface evolution processes with Lévi diffusion [3], the analysis of models for phase transition problems with long range interactions [15].

2 Fractional Isoperimetric Problems

The isoperimetric problem is one of the most studied and well known problems in mathematics. In the Euclidean space $\mathbb{R}^n$, it can be formulated as follows:

Problem 1 Prove that the solution to

$$\min_{E\subset\mathbb{R}^n} \frac{P(E)^{\frac{1}{n-1}}}{|E|^{\frac{1}{n}}} \tag{2.1}$$

is given by the euclidean ball.

Remark 3 The quantity in (2.1) is invariant under rigid transformations of $\mathbb{R}^n$ because both the perimeter and the volume are; it is also invariant under scaling, because of the choice of the exponents.

We observe here that Problem 1 is equivalent to the following

Problem 2 Prove that the solution to

$$\min_{E\subset\mathbb{R}^n, |E|=1} P(E)$$

is given by the euclidean ball.

The first proofs that the ball is the solution of Problem 1 were given in the nineteenth century. Nowadays, many strategies to prove this result are known. Among them, we can mention symmetrization, mass transportation and variational methods. In particular, the Steiner symmetrization technique consists in the following strategy: take a set E and fix a direction $\vartheta \in \mathbb{S}^{n-1}$. Without loss of generality, we can assume $\vartheta = e_1$. Then, for any $r \in \mathbb{R}$ define the (possibly empty) slice $E_r := \{(r, y) \in E\}$. Next, set $E_r^* = \{r\} \times B_{E_r}(0)$, where $B_{E_r}(0) \subset \mathbb{R}^{n-1}$ is the unique ball centered at $0 \in \mathbb{R}^{n-1}$ such that $|B_{E_r}(0)| = |E_r|$. Finally, define $E^* := \{(r, E_r^*) : r \in \mathbb{R}\}$. By Fubini theorem, we have that $|E^*| = |E|$, and one can notice that $P(E^*) \leq P(E)$. Therefore, with this symmetrization, we improved the isoperimetric ratio, namely

$$\frac{P(E^*)^{\frac{1}{n-1}}}{|E^*|^{\frac{1}{n}}} \leq \frac{P(E)^{\frac{1}{n-1}}}{|E|^{\frac{1}{n}}}.$$

We can then iterate the same argument along any direction $\vartheta \in \mathbb{S}^{n-1}$, until we reach the only set in $\mathbb{R}^n$ which is invariant for the symmetrization along any direction, namely the ball.

At the present day, the isoperimetric inequality is deeply understood, even in more sophisticated and sharp formulations, as the following result shows:

Theorem 3 (Sharp quantitative isoperimetric inequality (cf. [12, 9])) *There exists a constant $c_n > 0$ such that, for any $E \subset \mathbb{R}^n$ it holds*

$$\frac{P(E)^{\frac{1}{n-1}}}{|E|^{\frac{1}{n}}} \geq \frac{P(B)^{\frac{1}{n-1}}}{|B|^{\frac{1}{n}}} + c_n\, A(E)^2,$$

where

$$A(E) := \inf_{x \in \mathbb{R}^n} \left\{ \frac{|E \triangle B_E(x)|}{|E|} \right\}$$

and $B_E(x)$ denotes the ball centered at $x \in \mathbb{R}^n$ having the same volume of E.

Remark 4 The term $A(E)$ is called "Fraenkel asymmetry", and it represents a *discrepancy* from the optimal configuration i.e., the ball. Indeed, the term $A(E)$ is a way to measure how far a certain set is from being a ball: notice that $A(E) = 0$ if and only if E is an euclidean ball. We also remark that the exponent 2 over the term $A(E)$ cannot be improved.

Keeping in mind Theorem 1, one may investigate a fractional analogue of Problem 1:

Problem 3 Prove that the solution to

$$\min_{E \subset \mathbb{R}^n} \frac{P_\alpha(E)^{\frac{1}{n-\alpha}}}{|E|^{\frac{1}{n}}} \tag{2.2}$$

is given by the euclidean ball.

Remark 5 Even in this case, as in Remark 3, the quantity in (2.2) is invariant under rigid transformations and under scaling.

In order to prove that the euclidean ball actually solves Problem 3, a symmetrization technique can still be used, but it has to rely on a slightly more sophisticated tool than the Steiner symmetrization: in order to present it in Theorem 4, let us first give the following

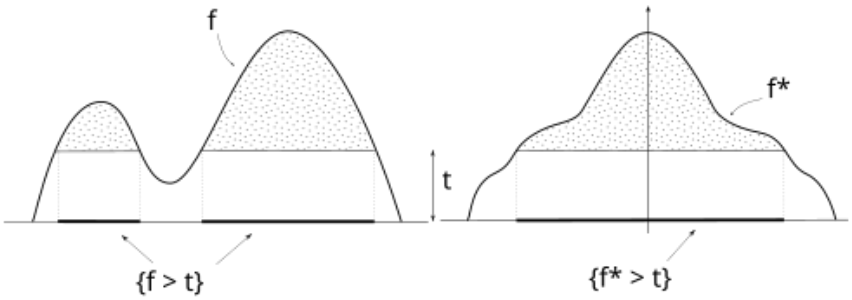
Definition 2 For any function $f\colon \mathbb{R}^n \to \mathbb{R}_{\geq 0}$. Let $\{f \geq t\} = \{y \in \mathbb{R}^n \colon f(y) \geq t\}$, and for any set $E \subset \mathbb{R}^n$, let χ_E denote its characteristic function. We define the symmetric decreasing rearrangement of f to be

$$f^*(x) \coloneqq \int_0^{+\infty} \chi_{\{f\geq t\}^*}(x)\, dt \quad \text{for any } x \in \mathbb{R}^n.$$

Remark 6 Definition 2 becomes meaningful if one keeps in mind the so called *layer cake formula*: for any function $f\colon \mathbb{R}^n \to \mathbb{R}_{\geq 0}$, it holds

$$f(x) = \int_0^{+\infty} \chi_{\{f\geq t\}}(x)\, dt \quad \text{for any } x \in \mathbb{R}^n.$$

See here below a figure in which the operation of symmetric decreasing rearrangement is visualized: we can rephrase this definition by saying that, in order to generate the symmetric decreasing rearrangement of a function f, we should replace every level set of f with the ball centered at the origin having the same volume, and consider the function f^* which has all these balls as level sets.



We are now ready to present the symmetrization tool needed to solve Problem 3:

Theorem 4 (Riesz Rearrangement Inequality) *Let $f, g, h\colon \mathbb{R}^n \to \mathbb{R}_{\geq 0}$, then it holds*

$$\iint_{\mathbb{R}^n\times\mathbb{R}^n} f(x)\, g(x-y)\, h(y)\, dx\, dy \leq \iint_{\mathbb{R}^n\times\mathbb{R}^n} f^*(x)\, g^*(x-y)\, h^*(y)\, dx\, dy, \tag{2.3}$$

where f^*, g^*, h^* are the symmetric decreasing rearrangements of f, g, h respectively.

With Theorem 4 in hand, one can solve Problem 3. Indeed, take a set $E \subset \mathbb{R}^n$, and define in (2.3) $f = h = \chi_E$, $g = \min\{M, |\cdot|^{-n-\alpha}\}$, where $M > 0$ is any fixed number.

Notice that g is symmetric decreasing, and hence $g^* = g$, whereas for $f = h$ it holds $f^* = h^* = (\chi_E)^* = \chi_{E^*}$, where E^* is the ball having the same volume as E. By Theorem 4, we then have

$$\begin{aligned}
&\iint_{\mathbb{R}^n\times\mathbb{R}^n} \frac{|\chi_E(x)-\chi_E(y)|^2}{\max\{|x-y|^{n+\alpha}, M^{-1}\}}\,dx\,dy\\
&= 2\iint_{\mathbb{R}^n\times\mathbb{R}^n} \frac{|\chi_E(x)|^2}{\max\{|x-y|^{n+\alpha}, M^{-1}\}}\,dx\,dy - 2\iint_{\mathbb{R}^n\times\mathbb{R}^n} \frac{\chi_E(x)\chi_E(y)}{\max\{|x-y|^{n+\alpha}, M^{-1}\}}\,dx\,dy\\
&= 2\int_{\mathbb{R}^n} \frac{|\chi_E(x)|^2}{k_M}\,dx - 2\iint_{\mathbb{R}^n\times\mathbb{R}^n} \frac{\chi_E(x)\chi_E(y)}{\max\{|x-y|^{n+\alpha}, M^{-1}\}}\,dx\,dy, =: I_1 + I_2
\end{aligned}$$

where $k_M = \int_{\mathbb{R}^n} g(x-y)\,dy = \int_{\mathbb{R}^n} g(y)\,dy$. Applying Fubini in I_2 and Theorem 4 in I_2, we deduce

$$\begin{aligned}
&\iint_{\mathbb{R}^n\times\mathbb{R}^n} \frac{|\chi_E(x)-\chi_E(y)|^2}{\max\{|x-y|^{n+\alpha}, M^{-1}\}}\,dx\,dy\\
&\geq 2\iint_{\mathbb{R}^n} \frac{|\chi_{E^*}(x)|^2}{k_M}\,dx - 2\iint_{\mathbb{R}^n\times\mathbb{R}^n} \frac{\chi_{E^*}(x)\chi_{E^*}(y)}{\max\{|x-y|^{n+\alpha}, M^{-1}\}}\,dx\,dy\\
&= \iint_{\mathbb{R}^n\times\mathbb{R}^n} \frac{|\chi_{E^*}(x)-\chi_{E^*}(y)|^2}{\max\{|x-y|^{n+\alpha}, M^{-1}\}}\,dx\,dy.
\end{aligned}$$

Noting that $|\chi_F(x)-\chi_F(y)|^2 = |\chi_F(x)-\chi_F(y)|$ for any set $F\subset\mathbb{R}^n$ and any $x, y\in\mathbb{R}^n$, and applying the monotone convergence theorem as $M\to+\infty$, we obtain

$$P_\alpha(E) = \frac{1}{2}\iint_{\mathbb{R}^n\times\mathbb{R}^n} \frac{|\chi_E(x)-\chi_E(y)|}{|x-y|^{n+\alpha}}\,dx\,dy \geq \frac{1}{2}\iint_{\mathbb{R}^n\times\mathbb{R}^n} \frac{|\chi_{E^*}(x)-\chi_{E^*}(y)|}{|x-y|^{n+\alpha}}\,dx\,dy = P_\alpha(E^*).$$

Since $|E^*| = |E|$, we proved Problem 3.

As in the case of Problem 1, we have an analogue of Theorem 3:

Theorem 5 (Sharp Quantitative Fractional Isoperimetric Inequality (cf. [10])) *There exists a constant $c_{\alpha,n} > 0$ such that, for any $E\subset\mathbb{R}^n$ it holds*

$$\frac{P_\alpha(E)^{\frac{1}{n-\alpha}}}{|E|^{\frac{1}{n}}} \geq \frac{P_\alpha(B)^{\frac{1}{n-\alpha}}}{|B|^{\frac{1}{n}}} + c_{n,\alpha}\,A(E)^2.$$

2.1 Isoperimetric inequality between two fractional perimeters

We now introduce the problem which will be object of Section 3. This is namely a fractional isoperimetric problem that arises as a modification of Problem 3 where, in view of Theorem 2, we replace the volume in with another fractional perimeter:

Problem 4 Let $0\leq s<t\leq 1$. Prove that the solution to

$$\min_{E\subset\mathbb{R}^n} \frac{P_t(E)^{\frac{1}{n-t}}}{P_s(E)^{\frac{1}{n-s}}} \tag{2.4}$$

is given by the euclidean ball.

Remark 7 Even in this case, the above quantity is scale invariant, due to the choice of exponents.

We stress that we improperly admitted the use of $s = 0, t = 1$: this has to be understood as an abuse of notation in the sense that $P_0(E) = |E|$, $P_1(E) = P(E)$.

The interest in Problem 4 is twofold. On the one hand, this problem provides the most natural generalization to Problem 3, which is recovered when the parameter $s \to 0$. Problem 3, in turn, was a generalization of Problem 1, so that the full range of fractional isoperimetric inequalities is covered by Problem 4. On the other hand, a solution to this problem cannot be achieved with symmetrizations techniques: indeed, as we saw earlier, the Riesz rearrangement provides a sharp tool to decrease the fractional perimeter of a set while keeping fixed its volume. In (2.4), however, both the numerator and denominator would decrease under this symmetrization, not revealing any monotonicity.

Problem 4 was first investigated in [8], where the authors proved the following result, concerning existence and regularity of solutions:

Theorem 6 *For any $0 \leq s < t \leq 1$, Problem 4 admits a solution E. Moreover, E is bounded, and has boundary of class $C^{1,\beta}$ for some $\beta = \beta(n, t-s)$ outside a closed singular set of dimension $n-3$ (respectively, $n-8$ if $t = 1$).*

3 Main result

The conjecture exposed in Problem 4 is still open. Yet, in our work [1], we prove a partial result that supports it. More precisely, we show that the euclidean ball is a local stable minimizer for the isoperimetric ratio in (2.4). This is equivalent to say that, among all sets which are near (in a suitable sense which we are going to explain soon) the ball, the ball minimizes the quantity in (2.4).

This kind of result (i.e., stability for perturbations of the ball) was proved also for Problem 1 by Fuglede [11], where he referred to these sets as *nearly spherical*. Actually, when proving quantitative isoperimetric inequalities, the typical ingredients are the following (see for example [11], [12]):

i) a *strict qualitative inequality*, showing that the ball is the unique minimizer;

ii) a reduction to the small asymmetry regime, establishing that it is sufficient to prove a quantitative isoperimetric inequality for sets whose L^1-distance from a ball is small;

iii) an aforementioned *Fuglede-type stability result*, showing the quantitative isoperimetric inequality in the class of "nearly spherical" sets.

From this, and also applying a selection principle in the spirit of [5], one might get the global quantitative inequality in the form of Theorem 3. Our result (see Theorem 7 below) represents step (*iii*) in the above program.

Before giving the precise statement of the result, together with a sketch of the proof, we need the following

Definition 3 Let $B(y,r)$ denote the Euclidean ball centred at $y \in \mathbb{R}^n$, with radius $r > 0$, and let B denote the unit ball at the origin. We say that E is a *graph over the ball* $B(y,r)$ if there exists a C^1 function $u\colon \partial B \to \mathbb{R}$ such that

$$E := \{y + \lambda(1 + u(x))x\colon \lambda \in [0,r], x \in \partial B\}.$$

The main result in [1] is the following:

Theorem 7 *Let $0 \le s < t \le 1$ and $n \ge 2$. Then there exists $\varepsilon_1 > 0$, depending only on s, t, and n, such that, if E is a graph over any ball via a function $u \in C^1(\partial B)$, with $\|u\|_{C^1} \le \varepsilon_1$, then*

$$\frac{P_t(E)^{\frac{1}{n-t}}}{P_s(E)^{\frac{1}{n-s}}} \ge \frac{P_t(B)^{\frac{1}{n-t}}}{P_s(B)^{\frac{1}{n-s}}},$$

and equality holds if and only if E is a ball.

Actually, Theorem 7 is the direct consequence of the following result, which can be seen as a local version of a sharp quantitative isoperimetric inequality in the spirit of Theorem 3 and Theorem 5

Theorem 8 *Let $0 \le s < t \le 1$ and $n \ge 2$. Then, there exist constants $\varepsilon_0, c_0 > 0$, depending only on s, t, and n, such that the following holds. Let E be a set, and let $B(y,r)$ be the ball having the same barycenter and volume as E. If E is a graph over $B(y,r)$ via a function u satisfying $\|u\|_{C^1} \le \varepsilon_0$, then*

$$\frac{P_t(E)^{\frac{1}{n-t}}}{P_s(E)^{\frac{1}{n-s}}} \ge \frac{P_t(B)^{\frac{1}{n-t}}}{P_s(B)^{\frac{1}{n-s}}} + c_0 \|u\|^2_{H^{\frac{1+t}{2}}},$$

where

$$\|u\|^2_{H^{\frac{1+t}{2}}(\partial B)} = \|u\|^2_{L^2(\partial B)} + \iint_{\partial B \times \partial B} \frac{|u(x) - u(y)|^2}{|x-y|^{n+t}}\, d\sigma_x\, d\sigma_y.$$

Here the term $\|u\|_{H^{\frac{1+t}{2}}}$ plays the role of the Fraenkel asymmetry of Theorem 3 and Theorem 5. Here, as stressed in Remark 4 for the analogous case, the exponent 2 is sharp.

3.1 Strategy of the proof

The result can be obtained through the following steps

Step 1. As described in [10] we can rewrite the functional $P_t(E)^{\frac{1}{n-t}} P_s(E)^{-\frac{1}{n-s}}$ as a functional $\mathscr{F}$ defined over the class of bounded functions on the unit sphere $\mathscr{C} = \{u\colon \partial B \to$

$\mathbb{R}, \|u\|_{L^\infty} \leq 1/2\}$. This allows to compute the variation of the isoperimetric ratio along perturbation in the class $\mathscr{C}$. It is not difficult to see that

$$\delta\mathscr{F}(0)[u] :- \frac{d}{d\varepsilon}\Big|_{\varepsilon=0} \mathscr{F}(\varepsilon u) = 0 \quad \text{for every } u \in \mathscr{C}, \tag{3.1}$$

and thus that the ball is a critical point for the isoperimetric ratio described in (2.4).

Step 2. We develop a second order analysis, showing that the functional $\mathscr{F}$ is coercive at the origin with respect to the norm $\|u\|_{H^{\frac{1+t}{2}}}$: there exists $c_0 = c_0(n, s, t) > 0$ such that

$$\delta^2\mathscr{F}(0)[u, u] \geq c_0 \|u\|^2_{H^{\frac{1+t}{2}}} \quad \text{for any } u \in \mathscr{C}. \tag{3.2}$$

This can be deduced computing explicitly $\delta^2\mathscr{F}(0)[u, u]$ and expressing this value in terms of the Fourier coefficients of the harmonic spherical decomposition of u (cf. [10]).

Step 3. Apart from coercivity, we also need a continuity estimate at the second order for $\mathscr{F}$: for any $\varepsilon_0 > 0$, there exists $c_1 = c_1(n, s, t) > 0$ such that

$$|\delta^2\mathscr{F}(u)[u, u] - \delta^2\mathscr{F}(0)[u, u]| \leq c_1 \|u\|_{C^1} \|u\|^2_{H^{\frac{1+t}{2}}} \quad \text{for any } u \in \mathscr{C} \text{ with } \|u\|_{C^1} \leq \varepsilon_0. \tag{3.3}$$

can be obtained after a careful analysis and symmetrization of the expression appearing on the left hand side. In particular, we rely on the symmetric structure of the kernel ruling the interaction in the definition of fractional perimeter, namely $|\cdot|^{-n-\alpha}$.

Step 4. Finally, collecting (3.1), (3.2) and (3.3) and performing a Taylor expansion on for the function $f\colon [0, 1] \to \mathbb{R}$, where $f(t) = \mathscr{F}(tu)$ for a fixed $u \in \mathscr{C}$, we obtain

$$\begin{aligned}
\mathscr{F}(u) = f(1) &= f(0) + f'(0) + \int_0^1 f''(t)(1-t)\,dt \\
&= \mathscr{F}(0) + \int_0^1 \delta^2\mathscr{F}(tu)[u, u](1-t)\,dt \\
&\geq \mathscr{F}(0) + \int_0^1 \delta^2\mathscr{F}(tu)[tu, tu]\frac{(1-t)}{t^2}\,dt - \int_0^1 \delta^2\mathscr{F}(0)[tu, tu]\frac{(1-t)}{t^2}\,dt \\
&+ \int_0^1 \delta^2\mathscr{F}(0)[tu, tu]\frac{(1-t)}{t^2}\,dt \\
&\geq \mathscr{F}(0) - \int_0^1 \left|\delta^2\mathscr{F}(tu)[tu, tu] - \delta^2\mathscr{F}(0)[tu, tu]\right| \frac{1-t}{t^2} + \int_0^1 \delta^2\mathscr{F}(0)[tu, tu]\frac{(1-t)}{t^2}\,dt \\
&\geq \mathscr{F}(0) - c_1\|u\|_{C^1}\|u\|^2_{H^{\frac{1+t}{2}}} + c_0\|u\|^2_{H^{\frac{1+t}{2}}} \geq \mathscr{F}(0),
\end{aligned}$$

provided ε_0 is small enough.

References

[1] Giovanni Alberti et al., *Stability of the ball in isoperimetric inequalities between two fractional perimeters*. cvgmt preprint. 2026. url: http://cvgmt.sns.it/paper/7715/.

[2] Jean Bourgain, Haim Brezis, and Petru Mironescu, *Another look at Sobolev spaces*. In: (2001).

[3] Luis Caffarelli and Luis Silvestre, *An extension problem related to the fractional Laplacian*. In: Comm. Partial Differential Equations 32.7-9 (2007), pp. 1245–1260. issn: 0360-5302,1532-4133. doi: 10.1080/03605300600987306. url: https://doi.org/10.1080/03605300600987306.

[4] Luis A Caffarelli, J-M Roquejoffre, and Ovidiu Savin, *Nonlocal minimal surfaces*. In: Comm. Pure Appl. Math. 63 (2010), pp. 1111–1144.

[5] Marco Cicalese and Gian Paolo Leonardi, *A selection principle for the sharp quantitative isoperimetric inequality*. In: Arch. Ration. Mech. Anal. 206.2 (2012), pp. 617–643. issn: 0003-9527,1432-0673. doi: 10.1007/s00205- 012- 0544- 1. url: https://doi.org/10.1007/s00205-012-0544-1.

[6] Eleonora Cinti, Joaquim Serra, and Enrico Valdinoci, *Quantitative flatness results and BV-estimates for stable nonlocal minimal surfaces*. In: Journal of Differential Geometry 112.3 (2019), pp. 447–504.

[7] Juan Dávila, *On an open question about functions of bounded variation*. In: Calculus of Variations and Partial Differential Equations 15.4 (2002), p. 519.

[8] Agnese Di Castro et al, *Nonlocal quantitative isoperimetric inequalities*. In: Calculus of Variations and Partial Differential Equations 54.3 (2015), pp. 2421–2464.

[9] Alessio Figalli, Francesco Maggi, and Aldo Pratelli, *A mass transportation approach to quantitative isoperimetric inequalities*. In: Inventiones Mathematicae 182.1 (2010), pp. 167–211.

[10] Alessio Figalli et al., *Isoperimetry and stability properties of balls with respect to nonlocal energies*. In: Communications in Mathematical Physics 336 (2015), pp. 441–507.

[11] Bent Fuglede, *Stability in the isoperimetric problem for convex or nearly spherical domains in* $\mathbb{R}^n$. In: Transactions of the American Mathematical Society 314.2 (1989), pp. 619–638.

[12] Nicola Fusco, Francesco Maggi, and Aldo Pratelli, *The sharp quantitative isoperimetric inequality*. In: Annals of mathematics (2008), pp. 941–980.

[13] Francesco Maggi, "Sets of finite perimeter and geometric variational problems: an introduction to Geometric Measure Theory". Cambridge University Press, 2012.

[14] Vladimir Maz'ya and Tatyana Shaposhnikova, *On the Bourgain, Brezis, and Mironescu theorem concerning limiting embeddings of fractional Sobolev spaces*. In: Journal of Functional Analysis 195.2 (2002), pp. 230–238.

[15] Otmar Scherzer et al., "Variational methods in imaging". Vol. 167. Springer, 2009.

Methods for Complex Network Analysis

Matteo Bergamaschi (*)

Abstract. In this note, we present an overview of modern methods for complex network analysis, with particular emphasis on propagation processes and optimization problems on graphs. Starting from the mathematical foundations of graph theory, we introduce classical propagation models and formulate the Influence Maximization (IM) problem as a mixed-integer optimization program under the General Information Propagation (GIP) model. We then describe the Network-aware Direct Search (NaDS) method, an efficient derivative-free optimization algorithm that exploits the graph structure to reduce the neighborhood explored at each iteration and proves that it converges to a d-local maximum. We further adapt NaDS to the Technology Diffusion Problem via binary search on the seed-set size. Finally, we discuss an application to modelling and containing financial distress propagation among Italian municipalities.

1 Introduction

Complex systems arising in social sciences, biology, economics, and technology are naturally represented as networks. The mathematical study of such structures is based on graph theory.

Definition 1 A *graph* is a pair

$$\mathcal{G}(\mathcal{N}, \mathcal{E}),$$

where $\mathcal{N} = \{1, \dots, n\}$ is a finite set of vertices (nodes) and $\mathcal{E} \subseteq \mathcal{N} \times \mathcal{N}$ is a set of edges connecting pairs of vertices.

Graphs may be classified according to the nature of their edges.

Definition 2 A graph is as follows:

- *undirected* if $\{i, j\} \in \mathcal{E}$ implies $\{j, i\} \in \mathcal{E}$;
- *directed* if edges possess an orientation, that is, $(i, j) \in \mathcal{E}$ does not necessarily imply $(j, i) \in \mathcal{E}$;
- *weighted* if each edge (i, j) is associated with a weight $w_{ij} > 0$ representing intensity, capacity, or interaction strength.

(*)Ph.D. course, Università di Padova, Dip. Matematica, via Trieste 63, I-35121 Padova, Italy. E-mail: bergamas@math.unipd.it . Seminar held on 23 April 2026.

Definition 3 Let $v \in \mathcal{N}$ be a node of a graph $\mathcal{G}(\mathcal{N}, \mathcal{E})$. The *neighbourhood* of v is

$$N(v) = \{u \in \mathcal{N} : (u, v) \in \mathcal{E}\},$$

and the *degree* of v is

$$d(v) = |N(v)|.$$

In directed graphs, one distinguishes the in-degree $d^{\text{in}}(v)$ and the out-degree $d^{\text{out}}(v)$.

Definition 4 The *distance of* the shortest path between two nodes $i, j \in \mathcal{N}$ is the minimum number of edges on a path connecting them and is denoted by $d(i, j)$. The *diameter* of the graph is

$$\max_{i,j \in \mathcal{N}} d(i, j).$$

A graph can also be represented algebraically through its adjacency matrix.

Definition 5 The *adjacency matrix* of a graph $\mathcal{G}(\mathcal{N}, \mathcal{E})$ is the matrix $A \in \{0, 1\}^{n \times n}$ defined by

$$A_{ij} = \begin{cases} 1 & \text{if } (i, j) \in \mathcal{E}, \\ 0 & \text{otherwise.} \end{cases}$$

For weighted graphs, the binary entry is replaced by the corresponding edge weight w_{ij}.

Two fundamental questions motivate the analysis developed in this work:

1. how information, behaviors, or shocks propagate through a network;
2. which nodes are most influential in determining that propagation.

These questions naturally lead to the study of propagation dynamics and the Influence Maximization problem.

2 The Influence Maximisation Problem

Influence Maximisation (IM) is the combinatorial optimization problem of selecting a small *seed set* of initially activated nodes in order to maximize the total expected spread of influence through a network. First formalized in [5], it has applications in viral marketing, public-health campaigns, and information dissemination.

2.1 Propagation models

All propagation models share a common iterative structure: (i) start with a set A_0 of active nodes; (ii) apply a model-specific activation rule to determine A_{t+1} from A_t; (iii) repeat until no further activations occur.

Linear Threshold Model (LT). Each node v has a threshold $\theta_v \in [0,1]$. The node v activates at time $t+1$ if the weighted influence of its active neighbors crosses the threshold:

$$\sum_{u \in N^{\text{in}}(v) \cap A_t} w_{uv} \;\geq\; \theta_v, \qquad \text{where} \sum_{u \in N^{\text{in}}(v)} w_{uv} \leq 1. \tag{1}$$

The model captures social conformity: a node joins only when sufficient peer pressure accumulates.

Independent Cascade Model (IC). When a node u becomes active at time t, it attempts to activate each inactive out-neighbor v with probability p_{uv}, independently and at most once. The model captures probabilistic word-of-mouth spreading.

General Information Propagation (GIP) Model. Used throughout this work, the GIP model assigns a continuous activation level $x_j(t) \geq 0$ to each node v_j. The dynamics are governed by

$$x_j(t) = f_{j,t}\left[\sum_{i=1}^{n} W_{ij}\, x_i(t-1)\right], \qquad \forall\, t > 0, \tag{2}$$

where $f_{j,t}$ is a piecewise-linear clipping function with a time-varying lower bound $l_{j,t}$ and an upper bound $h_{j,t}$:

$$f_{j,t}(x) = \begin{cases} 0 & x < l_{j,t}, \\ x & l_{j,t} \leq x \leq h_{j,t}, \\ h_{j,t} & x \geq h_{j,t}. \end{cases} \tag{3}$$

The bounds decay geometrically as $l_{j,t} = (\theta_{l,j}\alpha)^t l_{j,0}$ and $h_{j,t} = \theta_{h,j}\theta_{l,j}^{t-1}\alpha^t h_{j,0}$, capturing saturation and decay effects. An efficient propagation algorithm exploits the sparse activation structure: only nodes adjacent to currently active nodes need to be evaluated at each step, avoiding a full sweep of all n nodes per iteration.

2.2 The optimization problem

Under the GIP model, the IM problem is formulated as the following mixed-integer program with budget $B \ll n$:

$$\max_{\mathbf{x},\mathbf{z}} \; s(\mathbf{x}) \quad \text{s.t.} \quad x_j \leq h_{j,0} z_j, \quad x_j \geq l_{j,0} z_j, \quad \sum_{j=1}^{n} z_j \leq B, \quad x_j \in \mathbb{R},\; z_j \in \{0,1\}, \tag{4}$$

where z_j is a binary seed-selection variable, x_j is the initial activation level, and $s(\mathbf{x})$ is the total discounted influence spread computed by the GIP propagation algorithm. The problem is NP-hard and the objective is a black-box function admitting no exploitable gradient.

3 Network-aware Direct Search (NaDS)

3.1 Classical solution methods

Two families of classical approaches were reviewed.

- **Greedy algorithms.** Iteratively add the node providing the largest marginal gain. Under submodularity, this yields a $(1 - 1/e)$-approximation [5], but each iteration requires expensive influence-spread simulations, making the approach prohibitively slow on large networks.
- **Centrality-based heuristics.** Select nodes with high degree, Katz centrality [9], k-core [6], or Collective Influence [8]. Fast, but no performance guarantee.

3.2 Direct search methods and CDS

Direct search methods are derivative-free optimization algorithms that evaluate the objective in a neighborhood of trial points around the current solution, updating whenever a sufficient improvement is found. The Customized Direct Search (CDS) method [9] was the first DS algorithm applied to IM. Explores the discrete neighborhood

$$N(z, d) = \{y \in \{0, 1\}^n : \|y - z\|_1 \leq d,\ \|y\|_0 = B\}. \tag{5}$$

The key limitation is that $|N(z, d)|$ grows at least linearly with n, making CDS impractical for large networks.

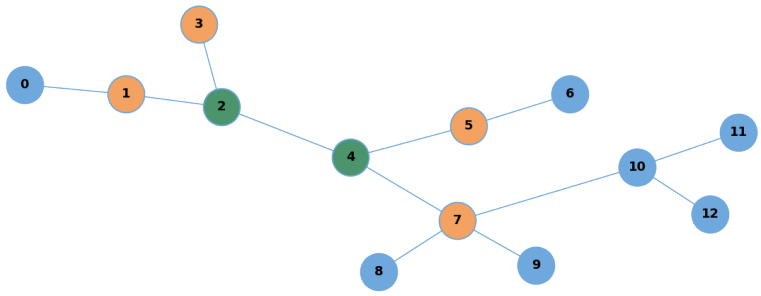


Figure 1: Simple example of how nodes are selected in CDS and NaDS. Given a set of activated nodes (in green), CDS selects potential swap nodes from all the remaining nodes of the graph (in blue and orange), while NaDS only considers the ones connected to the starting nodes (in orange).

3.3 The NaDS algorithm

Network-aware Direct Search [2] addresses the scalability bottleneck of CDS by restricting neighborhood exploration to a network-informed candidate set. Two empirical observations about large real-world networks motivate this:

- Mean node centrality does not change significantly as the network grows.

- The diameter does not grow significantly with the number of nodes (small-world property).

The candidate filtering set is defined as

$$C(z) = \{y \in \{0,1\}^n \mid y_i = 1 \Rightarrow \exists\, j : (i,j) \in \mathcal{E} \text{ or } z_i = 1\}. \tag{6}$$

Intuitively, $C(z)$ limits candidate seeds to nodes that are either already selected or directly connected to a selected node. For sparse networks, $|C(z)| \ll |N(z,d)|$.

The NaDS poll step proceeds in three phases.

(a) **Phase 1 – filtered search.** Evaluate s on $N(z^{(r)},2) \cap C(z^{(r)})$. If a point z with $s(z) > (1+\zeta)s(z^{(r)})$ is found, update immediately.

(b) **Phase 2 – full-neighborhood fallback.** If Phase 1 fails, expand to the full $N(z^{(r)},2)$ without the connectivity filter.

(c) **Phase 3 (optional) – radius expansion.** If $d < d_{\max}$, increase d and repeat.

The threshold ζ is progressively tightened ($\zeta \leftarrow \delta\zeta$) when improvement stalls; an optional global SEARCH step precedes the POLL step at each iteration.

Algorithm 1 Network-aware Direct Search (NaDS) Method

1: **Initialization:** Set $0 < \zeta, \delta < 1, d = 2, d_{max} > 1$. Let $z(0) \in \Omega^B$ be a suitably chosen starting point. Set iteration $r = 0$.
2: **SEARCH step (optional):** Evaluate s_{Ω^B} on a finite subset of trial points on the domain Ω^B, until a sufficiently improved point z is found, where $s_{\Omega^B}(z) > (1+\zeta)s_{\Omega^B}(z^{(r)})$, or all points have been exhausted. If an improved point is found, then the SEARCH step may terminate, skip the next POLL step and go directly to step 6.
3: **POLL step (Phase 1):** Evaluate s_{Ω^B} on the set $N(z^{(r)},d) \cap C(z^{(r)}) \subset \Omega^B$ as in (5) with distance d, until a sufficiently improved point z is found, where $s(\Omega^B)(z) > (1+\zeta)s(\Omega^B)(z^{(r)})$, or all points have been exhausted. $C(z^{(r)})$ is the set of points that are connected by an arc with $z^{(r)}$. If an improved point is found, then the POLL step may terminate, and go directly to step 6.
4: **POLL step (Phase 2):** Expand dynamically $N(z^{(r)},d) \cap C(z^{(r)})$ to $N(z^{(r)},d)$. If an improved point is found, then the POLL step may terminate, skip Phase 3, and go directly to step 6.
5: **POLL step (Phase 3 - Optional):** If $d < d_{max}$ then increase d.
6: **Termination check:** If an improvement is found, set $z^{(r+1)}$ as the improved solution, while decreasing $\zeta \leftarrow \delta\zeta$ if a sufficient improvement has not been found, increment $r \leftarrow r+1$, and go to step 2. Otherwise, output the solution $z^{(r)}$.

3.4 Theoretical guarantee

Definition 6 Given $z \in \{0,1\}^n$ and $d \in \mathbb{N}$, a point $z^* \in \{0,1\}^n$ is a *d-local maximum* of Problem 4 if $s(z^*) \geq s(z)$ for all $z \in N(z^*,d)$.

Theorem 1 *Let $\{z^{(r)}\}$ be the sequence of iterates generated by NaDS. Then the algorithm terminates at a d-local maximum for any ixed d.*

Proof. Let $\{z_k\}$ and $\{s_k\}$ be the sequences of solutions and reference values generated by NaDS. Since z_k is the output of NaDS, the **POLL step** phase ensures that every point z_t in the neighborhood $N(z_k, d)$ has been evaluated, and there is no point such that $s_t > s_k$. This guarantees that z_k is a d-local maximum. To prove that NaDS produces the sequence $\{z_k\}$ in finite time, we observe that $s_0 < s_1 < \cdots < s_k$, which implies that NaDS does not visit previously evaluated points twice. Furthermore, since the bounded domain Ω^B contains a finite number of points, it follows that the sequence $\{z_k\}$ must also be finite. □

3.5 Experimental evaluation

NaDS was benchmarked on six real-world collaboration and social networks (Arxiv Astro, Arxiv Gr-Qc, Arxiv Hep-Ph, LastFM-Asia, email-Enron, Facebook; ranging from 4,039 to 36,692 nodes) against CDS [9], Simple Greedy [5], Single Discount [3], Katz Centrality, K-core [6] and Collective Influence [8].

Dataset	Nodes	Edges
Arxiv Astro	18 772	198 110
Arxiv Gr-Qc	5242	14 496
Arxiv Hep-Ph	12 008	118 521
LastFM-Asia	7624	27 806
email-Enron	36 692	183 831
Facebook	4039	88 234

Table 1: Real-world networks used in the experiments.

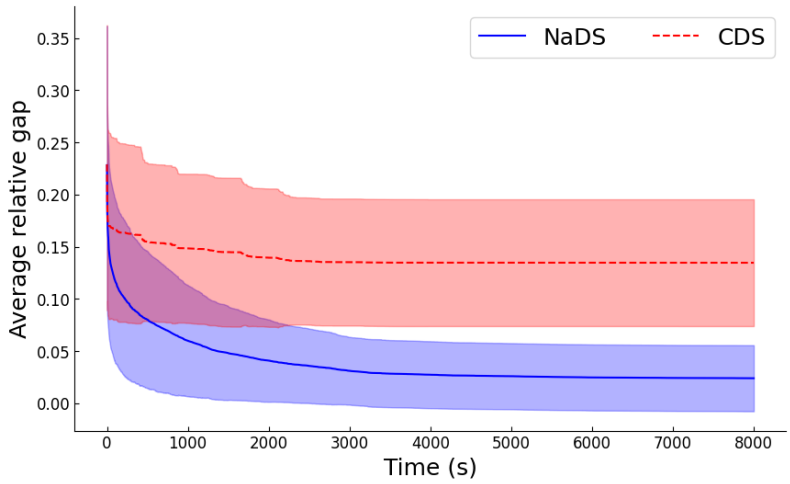


Figure 2: Average optimality gap (%) across all six networks and budget values B, comparing NaDS with CDS. Lower is better.

The GIP model was configured with $\theta_l = 2$ (requiring at least two active neighbors) and $\theta_h = 50$ (introducing behavioral heterogeneity); these settings make greedy methods less effective since the influence function loses submodularity under high-threshold dynamics. Two experimental protocols were considered: a single-start protocol initialized from a Single Discount heuristic solution, and a multi-start protocol with multiple near-heuristic starting points, both with time budgets scaled to network size. Results consistently showed that NaDS achieves lower optimality gaps than CDS and all non-DS baselines, with substantially reduced computation times on larger networks, validating the hypothesis that network-aware candidate filtering reduces expensive function evaluations without sacrificing solution quality.

4 Technology Diffusion Problem

4.1 Propagation model and problem formulation

The Technology Diffusion (TD) problem, introduced by Goldberg and Liu [4], uses a component-size-based activation rule. Each node u carries an integer threshold $\theta(u) \in \mathbb{N}$.

Cascade Rule. A node u *activates* when it is adjacent to a connected component of already-active nodes whose size exceeds $\theta(u)$: $|\mathcal{C}_{\text{active}}(u)| > \theta(u)$.

The TD optimization problem is to minimize the seed set size subject to the constraint that the resulting cascade activates the entire network:

$$\min_{\mathbf{z}} \|\mathbf{z}\|_0 \quad \text{s.t.} \quad \text{cascade}(\mathbf{z}) = \mathcal{N}, \quad z_j \in \{0,1\},\ \forall\, j \in [n]. \tag{7}$$

The problem is NP-hard; improved approximation algorithms for special graph classes are known [7].

4.2 Existing approaches

- **Integer programming (IP) formulations;** The diffusion process is encoded as a binary program. An approximate formulation has $O(n^2)$ variables and constraints; the exact formulation has $O(n^3)$. Optimal or small instances but entirely intractable at scale.
- **Centrality-based heuristics.** Nodes are added to the seed set in decreasing order of degree, betweenness, or closeness until full activation is achieved. Fast but without any performance guarantee.

4.3 NaDS adaptation via binary search

The key contribution [1] is an adaptation of NaDS to Problem 7 using binary search over the seed set size B:

1. Fix a budget B.
2. Run NaDS to find the best seed set of size B for a TD-flavored IM problem.

3. Check whether the seed set found triggers a full cascade.

4. If feasible, record the solution; if not, increase B and return to step 1.

The binary search on B requires only $O(\log n)$ NaDS runs to identify the minimum feasible seed set size. Experiments demonstrated that NaDS finds seed sets competitive with IP-optimal solutions on medium-scale instances and significantly outperforms centrality heuristics on large-scale ones, while remaining computationally tractable.

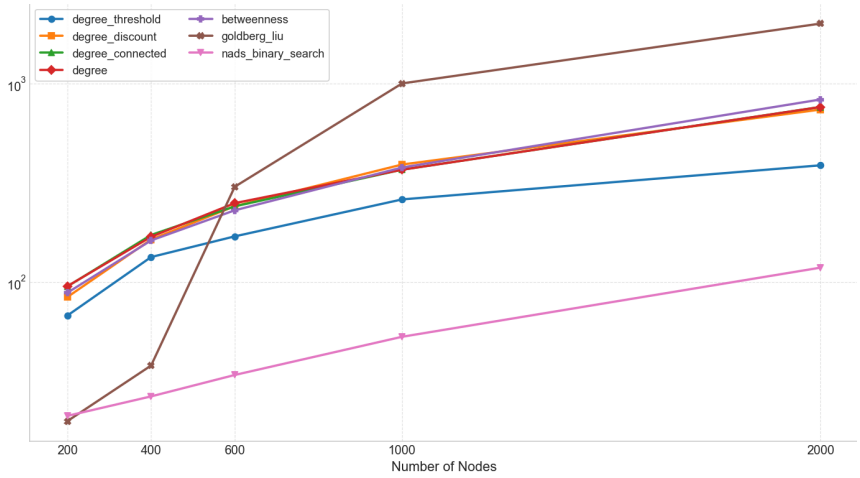


Figure 3: Mean budget B computed over synthetic graphs, comparing scalability as the number of nodes increases for NaDS against the $O(n^2)$ IP formulation of Goldberg and Liu and several heuristic methods.

5 Application: Public Finance and Municipal Financial Distress

5.1 Context

The final part of this note present an ongoing joint research project with Banca d'Italia on predicting and containing financial distress propagation among Italian municipalities. A first phase applies supervised machine learning to administrative, financial, and socio-economic features to produce early-warning indicators. The key conceptual advance is to move beyond node-level prediction towards a network model of distress propagation, viewing the containment problem through the IM lens.

5.2 Network construction

Three approaches to defining the weighted municipality graph were discussed: a distance-based graph with weights $w_{ij} = 1/\text{dist}(i,j)$ and a sparsifying threshold; Local Labour Systems (*Sistemi Locali del Lavoro*), Italian administrative zones defined by commuting flows; and economic interaction proxies based on mobility data, fiscal similarity, and shared service dependencies. A practical trade-off exists between graph density (richer information) and computational tractability (sparser graphs scale better for IM optimization).

5.3 Hyperedges and shared participations

An additional relational layer arises from shared corporate ownership: two municipalities may jointly hold a stake in the same firm, creating a shock-transmission channel not captured by pairwise edges. This structure naturally induces *hyperedges*, requiring an extension of the pairwise IM framework to hypergraph or bipartite formulations. This methodological extension is currently in progress.

6 Conclusions

This note presented a progression from graph-theoretic foundations to novel optimization methods and real-world applications. The NaDS algorithm exploits the small-world property of real-world networks to achieve scalable, guaranteed-quality solutions to the IM problem under the GIP model. Its adaptation to the Technology Diffusion Problem via binary search offers a practical alternative to intractable IP formulations.

Several directions for future work were identified: extending NaDS to dynamic networks, establishing approximation guarantees for the non-submodular GIP setting, completing the hypergraph extension for the public-finance application, and scaling the methodology to the full Italian municipality network of approximately 8,000 nodes.

References

[1] Matteo Bergamaschi et al., *A Direct Search approach for the Technology Diffusion Problem*. In: Ongoing work. 2026.

[2] Matteo Bergamaschi et al., *An efficient network-aware direct search method for influence maximization*. In: Journal of Complex Networks 13.6 (Dec. 2025), cnaf042. issn: 2051-1329. doi: 10.1093/comnet/cnaf042. eprint: `https://academic.oup.com/comnet/article-pdf/13/6/cnaf042/65284922/cnaf042.pdf`. url: `https://doi.org/10.1093/comnet/cnaf042`..

[3] Wei Chen, Yajun Wang, and Siyu Yang, *Efficient influence maximization in social networks*. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '09. New York, NY, USA: Association for Computing Machinery, 2009, 199–208. isbn: 9781605584959. doi: 10.1145/1557019.1557047. url: `https://doi.org/10.1145/1557019.1557047`.

[4] Sharon Goldberg and Zhenming Liu, *The Diffusion of Networking Technologies*. 2012. arXiv: 1202.2928 [cs.SI]. url: `https://arxiv.org/abs/1202.2928`.

[5] David Kempe, Jon Kleinberg, and Éva Tardos, *Maximizing the spread of influence through a social network*. In: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. 2003, pp. 137–146.

[6] Maksim Kitsak et al., *Identification of influential spreaders in complex networks*. In: Nature Physics 6.11 (Aug. 2010), 888–893. issn: 1745-248. doi: 10.1038/nphys1746. url: `http://dx.doi.org/10.1038/nphys1746`.

[7] Jochen Könemann, Sina Sadeghian, and Laura Sanità, *Better Approximation Algorithms for Technology Diffusion*. In: Algorithms - ESA 2013. Ed. by Hans L. Bodlaender and Giuseppe

F. Italiano. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 637–646. isbn: 978-3-642-40450-4.

[8] Flaviano Morone and Hernán A. Makse, *Influence maximization in complex networks through optimal percolation*. In: Nature 524.7563 (July 2015), 65–68. issn: 1476-4687. doi: 10.1038/nature14604. url: `http://dx.doi.org/10.1038/nature14604`.

[9] Yu Tian and Renaud Lambiotte, *Unifying Diffusion Models on Networks and Their Influence Maximisation*. In: CoRR abs/2112.01465 (2022). arXiv: 2112.01465. url: `https://arxiv.org/abs/2112.01465`.

A Fast Ride on the Turnpike

Nicolò De Bernardi [(*)]

Abstract. The term "turnpike" first appeared in econometrics in the 1950s, although the phenomenon had already been observed by Von Neumann in 1938. It describes a stability property of optimal trajectories: the optimal path between any two points is the one that moves rapidly from the initial condition toward a steady state (the "turnpike"), remains close to it for most of the time horizon, and adjusts to meet the terminal condition only near the final time, provided the endpoints are sufficiently far apart. This property has been established in various contexts, including control theory, dynamical systems, and Mean Field Games, under suitable monotonicity assumptions on the model.
In this note we will present some basic examples of turnpike properties, with the aim of highlighting the main structural features of the models that ensure their validity. Moreover, in the context of quadratic MFG, we will introduce a weaker monotonicity assumption that still guarantees the turnpike behavior toward local equilibria, along with an interpretation in terms of spectral properties of an operator encoded in the MFG system.

1 Introduction

In the English vocabulary, the word "turnpike" means "highway". Before being introduced in mathematics, the so-called turnpike phenomenon had first been observed in the context of econometrics, in particular by F. P. Ramsey (1928) and Von Neumann (1938) [9], to describe the optimal and most efficient way for capital accumulation. In 1958 this term appeared for the first time in [5], an article by the Nobel Prize winner Paul Samuelson, together with R. Dorfman and R. Solow. Roughly speaking, the idea behind it is that if the initial and terminal states are far enough apart, the optimal way to connect them can be divided into three steps:

(a) At the beginning, the optimal path moves rapidly from the initial condition in order to reach a steady state (the "turnpike");

(b) For most of the time horizon, the trajectory remains close to the turnpike;

(c) Near the final endpoint, it moves rapidly away from the turnpike in order to reach the terminal condition.

[(*)]Ph.D. course, Università di Padova, Dip. Matematica, via Trieste 63, I-35121 Padova, Italy. E-mail: debernar@math.unipd.it . Seminar held on 7 May 2026.

This means that, however far the endpoints are from the steady state, the optimal strategy is to stay for most of the time horizon close to it, provided the endpoints are far enough or the time horizon is long enough. A simple illustration of this phenomenon is provided in Figure 1.

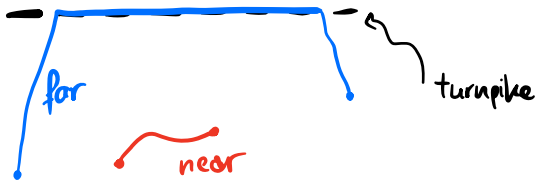
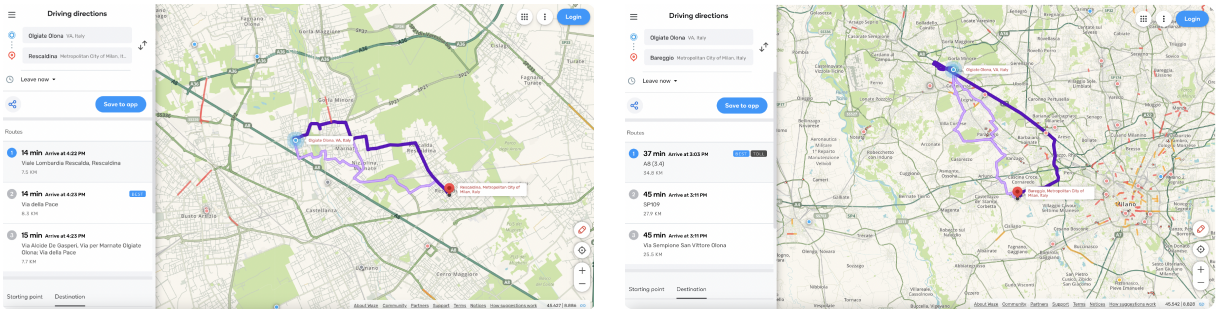


Figure 1: An illustration of the turnpike phenomenon.



(a) The endpoints are close enough: no turnpike.

(b) The endpoints are far enough apart: turnpike phenomenon occurs.

Figure 2: Two possible planned trips (from Waze).

This situation is also typical when planning a trip. The two maps displayed in Figure 2 represent the geographical area near Milan, Italy. In the upper map, it is possible to observe that, since the origin and the destination of the trip are close, the optimal route (actually all fastest routes) does not travel along the turnpike, but along parallel and slower streets. On the contrary, in the lower map, the fastest route is along the highway, because of the greater distance between the endpoints, even if it is not the shortest one and it involves a deviation from the origin to reach the turnpike.

1.1 An example

Let us provide an instructive example of such a property for a linear, two-dimensional dynamical system. Let $[x(t), y(t)] \in \mathbb{R}^2$. We define the system as

$$\begin{cases} \dot{x} = 2x + 3y \\ \dot{y} = -x - 2y \end{cases} \tag{1.1}$$

which can be equivalently rewritten as

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \underbrace{\begin{bmatrix} 2 & 3 \\ -1 & -2 \end{bmatrix}}_{M} \begin{bmatrix} x \\ y \end{bmatrix}. \tag{1.2}$$

Because of the linearity of the system and of the invertibility of M, it turns out that there exists just one critical point, located at $[\bar{x}, \bar{y}] = [0, 0]$. Moreover, it is straightforward to observe that the matrix M describing the system is Hamiltonian. Indeed, if

$$J = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix},$$

then JM is symmetric. It is known that for Hamiltonian matrices, if λ is an eigenvalue, also $-\lambda, \bar{\lambda}, -\bar{\lambda}$ are eigenvalues. In particular, in this example, the eigenvalues of M are 1 and -1, which implies that the critical point is of saddle type (see Figure 3).

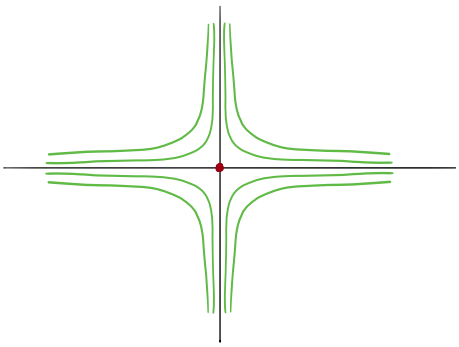


Figure 3: The origin $[0, 0]$ is an equilibrium of hyperbolic nature. This is due to the presence of two real eigenvalues, with opposite sign.

If one computes the explicit solution to (1.1) on a finite-time horizon $[0, T]$, the x component is given by

$$x(t) = \frac{x(0)e^{T-t} - x(T)e^{-t} + x(T)e^{t} - x(0)e^{-(T-t)}}{e^{T} - e^{-T}}, \tag{1.3}$$

whose plot is in Figure 4. Notice that for most of the time horizon $[0, T]$ the solution to the dynamical system remains close to the equilibrium and deviates from it near the endpoints to match with the initial / terminal conditions.

This exponential-type behavior occurs because the matrix M has only real eigenvalues and no purely imaginary eigenvalues, which would force the equilibrium to be of center type, hence without any stabilization toward it. This reflects the hyperbolic nature of the equilibrium, which is responsible for its exponential stability.

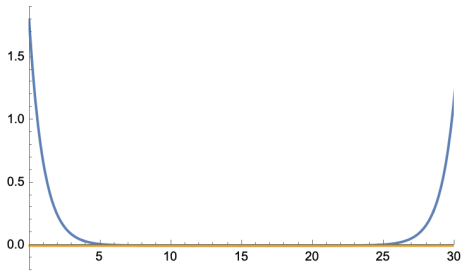


Figure 4: The blue curve represents the first component of the solution to (1.1), on $[0,T]$, while the yellow line stands for the critical point $\bar{x}=0$. Here $T=30$, $x(0)=1.8$, $x(30)=1.2$.

2 The MFG system and the local turnpike property

There has been an increasing interest in the study of turnpike properties in finite-dimensional optimal control problems, starting from the 60's (see for instance [7], [8]) whereas the analysis of turnpike phenomena is quite recent in the context of Mean Field Games (MFG) and Mean Field Control (MFC).
In this note, we study the long time behavior of solutions to the discounted Mean Field Games system of PDEs with quadratic Hamiltonian, which takes the following form:

$$\begin{cases} -\partial_t u - \Delta u + \frac{|Du|^2}{2} - f(x,m) + \delta u = 0, & \text{on } (0,T)\times\Omega \\ \partial_t m - \Delta m - \operatorname{div}(mDu) = 0, & \text{on } (0,T)\times\Omega \\ m(0,x) = m_0(x), \qquad u(T,x) = u_T(x), & \text{on } \Omega, \end{cases} \tag{2.1}$$

where $T>0$ is the time horizon, $f:\Omega\times\mathbb{R}\to\mathbb{R}$ is of class C^2 and has bounded first and second derivatives with respect to the second component, while $\delta\geq 0$ is a discount factor, which is assumed to be small. Moreover, Ω can be either the Euclidean space $\mathbb{R}^n$ or the flat torus $\mathbb{T}^n$ (in which case, data are periodic in the x variable); m_0, u_T are the initial distribution and the terminal cost, respectively, and they are assumed to be of class $C^{2,\alpha}$. This PDE system has been introduced in [6] to describe *Mean Field Nash equilibria*, that is the equilibrium configuration for a system of infinitely many indistinguishable players / agents, distributed with density $m(t,x)$, each of which controls his / her own state

$$\mathrm{d}X_s = \alpha_s \mathrm{d}s + \sqrt{2}\mathrm{d}B_s \qquad \in \Omega,$$

aiming at minimizing the cost functional

$$J(\alpha) = \mathbb{E}\int_0^T e^{-\delta s}\left(\frac{1}{2}|\alpha_s|^2 + f(X_s, m(s,X_s))\right)\mathrm{d}s + e^{-\delta T}u_T(X_T),$$

where u is the so-called value function of the typical agent. Roughly speaking, $u(t,x)$ represents how each agent should choose his / her control to play the optimal strategy starting at position x and time t. The first equation is a backward Hamilton-Jacobi-Bellman equation to describe the evolution of the value function u, while the second is a

forward Fokker-Planck equation for the evolution of the players' optimal distribution m in the Mean Field game.

Equivalently, the above problem can be rephrased as the system of optimality conditions for the Mean Field optimal control problem of the functional:

$$\mathcal{F}(w) = \int_0^T \int_\Omega e^{-\delta s} \left(\frac{1}{2} m(s,y)|w(s,y)|^2 + F(y, m(s,y)) \right) \mathrm{d}y \, \mathrm{d}s + \int_\Omega e^{-\delta T} u_T(y) \, \mathrm{d}y, \tag{2.2}$$

subject to the constraint

$$\partial_t m - \Delta m + \operatorname{div}(mw) = 0, \text{ on } (0,T) \times \Omega \qquad m(0) = m_0,$$

where $F(x,m) = \int_0^m f(x,\nu) \, \mathrm{d}\nu$.
Moreover, we consider a solution $(\bar{u}, \bar{m})$ to the corresponding stationary problem

$$\begin{cases} \Delta \bar{u} - \frac{|D\bar{u}|^2}{2} + f(x,\bar{m}) + \delta \bar{u} = 0, & \text{on } \Omega \\ \Delta \bar{m} + \operatorname{div}(\bar{m} D\bar{u}) = 0, & \text{on } \Omega \\ \int_\Omega \bar{m} = 1 \end{cases} \tag{2.3}$$

At least when $\delta = 0$, this system describes the optimality condition for

$$\overline{\mathcal{F}}(w) = \int_\Omega \frac{1}{2} m(y)|w(y)|^2 + F(y, m(y)) \mathrm{d}y$$

subject to

$$-\Delta m + \operatorname{div}(mw) = 0, \quad \text{on } \Omega \qquad \int_\Omega m(y) \mathrm{d}y = 1.$$

The main question we address is the following:

Under which conditions on $(\bar{u}, \bar{m})$ can we guarantee that it attracts dynamic equilibria (u,m) as $T \to +\infty$, at least if (u,m) is close enough to $(\bar{u}, \bar{m})$?

More precisely, we are looking for the assumptions such that there exist $\sigma_1, \sigma_2 > 0$ (possibly different) such that the following exponential-type turnpike inequality holds true:

$$\|m(t,\cdot) - \bar{m}(\cdot)\|_X + \|Du(t,\cdot) - D\bar{u}(\cdot)\|_Y \lesssim e^{-\sigma_1 t} + e^{-\sigma_2 (T-t)}, \tag{2.4}$$

where X, Y and their norms are to be precised.
In Mean Field Game theory, this property has been studied recently under some conditions which guarantee a global turnpike phenomenon. These assumptions take the form of monotonicity (or "smallness" of data), they ensure uniqueness of both dynamic and stationary equilibria, and the convergence of the former to the latter, or, at least on a finite time horizon, the turnpike behavior of the dynamic solution with respect to the stationary one. The main assumption is the so-called (strictly) *Lasry-Lions monotonicity*, introduced in [6] and used in [2], which takes the following form (for local couplings):

$$\int_\Omega f(x,m_1) - f(x,m_2)(m_1 - m_2) \, \mathrm{d}x \geq \gamma \int_\Omega (m_1 - m_2)^2 \, \mathrm{d}x, \tag{2.5}$$

for all $x \in \Omega$, for all m_1, m_2 probabilities with density, and for some $\gamma > 0$.
Our main aim is to address situations where such monotonicity fails. In particular, we cannot expect uniqueness of neither stationary equilibria nor dynamic solutions; we then need to focus on the local behavior around one stationary equilibrium and its stability properties.
The crucial assumptions we need the stationary state to satisfy are the following (note that $\bar{m}$ is everywhere positive on Ω):

(S)
- The identity $D\bar{m} = -\bar{m}D\bar{u}$ holds on Ω,
- $\bar{m}$ satisfies a Poincaré inequality with constant $C_P > 0$, i.e.

$$\int_\Omega g^2(x)\bar{m}(x)\mathrm{d}x - \left(\int_\Omega g(x)\bar{m}(x)\mathrm{d}x\right)^2 \leq C_P \int_\Omega |Dg|^2(x)\bar{m}(x)\mathrm{d}x \qquad \forall g \in W^{1,2}(\bar{m}\mathrm{d}x),$$

- there exists $\eta \in (0,1)$ such that

$$\int_\Omega f_m(x,\bar{m})\mu^2 + \bar{m}\left|D\left(\frac{\mu}{\bar{m}}\right)\right|^2 + \delta\frac{\mu^2}{\bar{m}}\,\mathrm{d}x \geq \eta\int_\Omega \bar{m}\left|D\left(\frac{\mu}{\bar{m}}\right)\right|^2\mathrm{d}x, \tag{2.6}$$

for all $\mu \in L^1(\Omega)\cap L^\infty(\Omega)$ such that $\int_\Omega \mu(x)\mathrm{d}x = 0$, $\mu/\bar{m} \in W^{1,2}(\bar{m}\mathrm{d}x)$, $\mu/\bar{m} \in L^\infty(\Omega)$.

The first hypothesis is expected to be true in our setting, where uniqueness of solutions to the (stationary) Fokker-Planck equation is guaranteed [1]. The second assumption is also quite easy to prove when $\Omega = \mathbb{T}^n$, while it is more delicate in the non-compact case, and it is strictly related to the decay of $\bar{m}$ as $x \to +\infty$. In the next Subsection, we elaborate more on the third hypothesis.

2.1 Local turnpike

Consider (u,m) and $(\bar{u},\bar{m})$ solutions to (2.1) and (2.3) respectively. Since we study local stability, our analysis will be of perturbative nature; hence, define

$$\mu := m - \bar{m} \qquad v := u - \bar{u},$$

which satisfy the following PDE system:

$$\begin{cases} -\partial_t v - \Delta v + \langle Dv, D\bar{u}\rangle + \delta v = f(x,\bar{m}+\mu) - f(x,\bar{m}) - \frac{|Dv|^2}{2}, & \text{on } (0,T)\times\Omega \\ \partial_t \mu - \Delta\mu - \mathrm{div}(\bar{m}Dv) - \mathrm{div}(\mu D\bar{u}) = \mathrm{div}(\mu Dv), & \text{on } (0,T)\times\Omega \\ \mu(0,\cdot) = \mu_0(\cdot) = m_0(\cdot) - \bar{m}(\cdot) \qquad v(T,\cdot) = v_T(\cdot) = u_T(\cdot) - \bar{u}(\cdot), & \text{on } \Omega. \end{cases} \tag{2.7}$$

From now on, for the sake of simplicity, we take $\delta = 0$. Consider now the linearization of (2.7):

$$\begin{cases} \partial_t \mu = \Delta\mu + \mathrm{div}(\bar{m}Dv) + \mathrm{div}(\mu D\bar{u}), \\ \partial_t v = -\Delta v + \langle Dv, D\bar{u}\rangle - f_m(x,\bar{m})\mu. \end{cases} \tag{2.8}$$

We are interested in establishing stability properties of the equilibrium state $(0,0)$. As in Subsection 1.1, we can write the above problem highlighting its encoded Hamiltonian structure.

$$\begin{bmatrix}\dot{\mu}\\ \dot{v}\end{bmatrix} = \begin{bmatrix} A & -BB^*\\ -Q & -A^*\end{bmatrix}\begin{bmatrix}\mu\\ v\end{bmatrix},$$

where

$$\begin{aligned} Ah &= \Delta h + \mathrm{div}(hD\bar{u}),\\ A^*h &= \Delta h - \langle Dh, D\bar{u}\rangle,\\ BB^*h &= -\mathrm{div}(\bar{m}Dh),\\ Qh &= f_m(x,\bar{m})h. \end{aligned}$$

Notice that the operator $h \mapsto -\mathrm{div}(\bar{m}Dh)$ is self-adjoint and positive, hence it can be written in the form BB^*, where $Bh = -\mathrm{div}(\sqrt{\bar{m}}h)$. Denote by M the matrix describing the linearized system: it is Hamiltonian because also Q is a self-adjoint operator and the diagonal blocks satisfy $A + ((-A)^*)^* = A - A = 0$.
As explained in Subsection 1.1, linear stability of $(0,0)$ and the turnpike behavior around it come from its hyperbolic nature: that is, one needs to prevent M from having purely imaginary eigenvalues, i.e. the following property must hold:

$$\mathrm{Ker}\,(M - i\omega I) = \{0\} \qquad \forall\, \omega \in \mathbb{R}, \omega \neq 0. \tag{2.9}$$

To do this, let $(\bar{\mu}, \bar{v})$ such that

$$\begin{bmatrix} A - i\omega I & -BB^*\\ -Q & -A^* - i\omega I\end{bmatrix}\begin{bmatrix}\bar{\mu}\\ \bar{v}\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix},$$

we have to verify that the only solution is $(\bar{\mu}, \bar{v}) = (0,0)$. From the first row, we derive that $\bar{v} = (BB^*)^{-1}(A - i\omega I)\bar{\mu}$, from which

$$\left\{Q + A^*(BB^*)^{-1}A + i\omega\left[(BB^*)^{-1}A - A^*(BB^*)^{-1}\right] + \omega^2(BB^*)^{-1}\right\}\bar{\mu} = 0.$$

Since $BB^* > 0$, to guarantee that the only solution is $\bar{\mu} = 0$ we may assume that

(i) $Q + A^*(BB^*)^{-1}A \geq 0$

(ii) $A^*(BB^*)^{-1} = (BB^*)^{-1}A$

Assumption (ii) is always satisfied within the quadratic Hamiltonian framework thanks to the validity of the relation $D\bar{m} = -\bar{m}D\bar{u}$, and it plays the role of a symmetry property, which may not hold in much more generality.
Condition (i) above can be rewritten (again exploiting $D\bar{m} = -\bar{m}D\bar{u}$) as follows

$$\int_\Omega f_m(x,\bar{m})\mu^2 + \bar{m}\left|D\left(\frac{\mu}{\bar{m}}\right)\right|^2 \mathrm{d}x \geq 0. \tag{2.10}$$

This assumption can be interpreted as a second-order condition ensuring that $(\bar{u}, \bar{m})$ is a minimizer (not just a critical point) of $\overline{\mathcal{F}}$.

A strict version of (2.10) is needed: in the next Section we will show why it is necessary. This strict inequality can be written in the operator form as

$$Q + A^*(BB^*)^{-1}A \geq \eta A^*(BB^*)^{-1}A, \tag{2.11}$$

for some $\eta \in (0,1)$. Recalling again the definition of the operators, (2.11) is precisely (2.6).

2.2 The Lyapunov function

To study the stability properties of solutions to (2.8), we need to find a Lyapunov function encoded in the system, that is a function $\Phi(s)$ such that there exists $\sigma > 0$ such that

$$-\dot{\Phi}(s) \geq \sigma|\Phi(s)| \qquad \forall s \in (0,T). \tag{2.12}$$

Indeed, this property is the core of the exponential stability: the above inequality can be split into

$$-\dot{\Phi}(s) \geq \sigma\Phi(s) \quad \text{and} \quad -\dot{\Phi}(s) \geq -\sigma\Phi(s) \qquad \forall s \in (0,T).$$

Integrating the first inequality on $(0,t)$ and the second on (t,T), for all $t \in (0,T)$, one gets

$$\Phi(T)e^{-\sigma(T-t)} \leq \Phi(t) \leq \Phi(0)e^{-\sigma t}, \tag{2.13}$$

recovering the exponential decay typical of the turnpike property.
In several papers about turnpike, such as [2], it is common to look at the evolution of the state / adjoint-state quantity:

$$\phi(t) := \int_\Omega \mu(t,x)v(t,x)\,\mathrm{d}x. \tag{2.14}$$

A direct computation shows that deriving an inequality of the form (2.12) requires the positivity of the operator Q defined in Section 2, or, equivalently, the Lasry–Lions monotonicity condition for f. Starting from the equation for μ, which can be rewritten as

$$-(BB^*)^{-1}\dot{\mu} = -(BB^*)^{-1}A\mu + v,$$

one can observe that it is more convenient to look at the evolution of

$$\begin{aligned}\Phi(t) :&= \int_\Omega \mu(t,x)v(t,x) - \mu(t,x)\cdot\left((BB^*)^{-1}A\mu(t,x)\right)\,\mathrm{d}x = \\ &= \int_\Omega \mu(t,x)v(t,x) + \frac{\mu^2(t,x)}{\bar{m}(x)}\,\mathrm{d}x.\end{aligned} \tag{2.15}$$

Assuming some integrability on μ, v and their derivatives, we have

$$\begin{aligned}-\dot{\Phi}(t) &= \int_\Omega (-\partial_t\mu)\cdot v - \mu\cdot(\partial_t v) - 2\frac{\mu}{\bar{m}}\partial_t\mu\,\mathrm{d}x = \int_\Omega (-\Delta\mu - \mathrm{div}(\bar{m}Dv) - \mathrm{div}(\mu D\bar{u}))\cdot v + \\ &+ \mu(\Delta v - \langle Dv, D\bar{u}\rangle + f_m(x,\bar{m})\mu) - 2\frac{\mu}{\bar{m}}\left(\Delta\mu + \mathrm{div}(\bar{m}Dv) + \mathrm{div}(\mu D\bar{u})\right)\,\mathrm{d}x = \\ &= \int_\Omega f_m(x,\bar{m})\mu^2 + \bar{m}|Dv|^2 + 2\bar{m}\langle Dv, D\left(\frac{\mu}{\bar{m}}\right)\rangle + 2\langle D\left(\frac{\mu}{\bar{m}}\right), D\mu + \mu D\bar{u}\rangle\,\mathrm{d}x.\end{aligned}$$

Starting from the equality $D\bar{m} = -\bar{m}D\bar{u}$, one can observe that $D\left(\frac{\mu}{\bar{m}}\right) = \frac{D\mu+\mu D\bar{u}}{\bar{m}}$, from which

$$\begin{aligned}
-\dot{\Phi}(t) &= \int_\Omega f_m(x,\bar{m})\mu^2 + \bar{m}|Dv|^2 + 2\bar{m}\langle Dv, D\left(\frac{\mu}{\bar{m}}\right)\rangle + 2\bar{m}\left|D\left(\frac{\mu}{\bar{m}}\right)\right|^2 \mathrm{d}x \\
&\overset{(2.6)}{\geq} \int_\Omega \bar{m}|Dv|^2 + 2\bar{m}\langle Dv, D\left(\frac{\mu}{\bar{m}}\right)\rangle + (1+\eta)\bar{m}\left|D\left(\frac{\mu}{\bar{m}}\right)\right|^2 \mathrm{d}x = \\
&= \int_\Omega \eta\bar{m}\left|D\left(\frac{\mu}{\bar{m}}\right)\right|^2 + \bar{m}\left|Dv + D\left(\frac{\mu}{\bar{m}}\right)\right|^2 \mathrm{d}x \geq \int_\Omega \frac{\eta}{4}\bar{m}|Dv|^2 + \frac{\eta}{4}\bar{m}\left|D\left(\frac{\mu}{\bar{m}}\right)\right|^2 \mathrm{d}x \\
&\geq \sigma \int_\Omega \left|\mu v + \frac{\mu^2}{\bar{m}}\right| \geq \sigma|\Phi(t)|,
\end{aligned}$$

where $\sigma > 0$ depends on η and on the Poincaré constant of $\bar{m}$, and the last inequalities are due to Young's and triangle inequalities.
The same arguments apply if we consider the nonlinear system (2.7). In this case, it suffices to deal with the nonlinear terms as perturbation of the linearized system, hence assuming some smallness assumptions on them.

2.3 The main results

Once the existence of a Lyapunov function has been established, by exploiting properly the equations for μ and v, it is possible to show the following *a priori* estimates, in the general discounted framework. These results can be found in [4].

Theorem 2.1 *Assume that (S) holds. Let (u,m) and $(\bar{u},\bar{m})$ be (classical) solutions to (2.1) and (2.3) respectively, such that*

$$\mu = m - \bar{m}, \qquad v = u - \bar{u}$$

satisfy some integrability conditions. Then, there exist $\bar{\varepsilon} > 0, \bar{\gamma} > 0, \sigma > 0$ (which depend only on $\bar{m}, \bar{u}, \eta, f$, but not on the time horizon T), such that the following is true for all $\gamma \leq \bar{\gamma}$, $\delta < \sigma$: if

$$\begin{aligned}
&\|m_0(\cdot) - \bar{m}(\cdot)\|_{L^\infty(\Omega)} + \|u(T,\cdot) - \bar{u}(\cdot)\|_{W^{1,\infty}(\Omega)} < \gamma, \\
&\left\|\frac{m_0(\cdot) - \bar{m}(\cdot)}{\sqrt{\bar{m}(\cdot)}}\right\|_{L^2(\Omega)} + \left\|\sqrt{\bar{m}(\cdot)}\,|Du(T,\cdot) - D\bar{u}(\cdot)|\right\|_{L^2(\Omega)} < \gamma, \\
&\|m(t,\cdot) - \bar{m}(\cdot)\|_{L^\infty(\Omega)} \leq \bar{\varepsilon}\left(e^{-\sigma_1 t} + e^{-\sigma_2(T-t)}\right) \qquad \forall t \in [0,T],
\end{aligned}$$

then

$$\|m(t,\cdot) - \bar{m}(\cdot)\|_{L^\infty(\Omega)} \leq \frac{\bar{\varepsilon}}{2}\left(e^{-\sigma_1 t} + e^{-\sigma_2(T-t)}\right) \qquad \forall t \in [0,T],$$

where $\sigma_1 = \frac{\sigma-\delta}{n+1}$ and $\sigma_2 = \frac{\sigma+\delta}{n+1}$.

The previous Theorem is rather technical and it is needed to apply a *Leray-Schauder* fixed point argument to show the existence (when $\Omega = \mathbb{T}^n$, to avoid non-compactness issues)

of *at least* one dynamic solution satisfying the (local) exponential turnpike property around $(\bar{u}, \bar{m})$, provided that the initial / terminal conditions are close enough to the stationary state. This result is contained in the following

Theorem 2.2 *Let $\Omega = \mathbb{T}^n$. Suppose that $(\bar{u}, \bar{m})$ is a (classical) solution to (2.3) and that (S) holds. Then, there exist constants $\bar{\varepsilon} > 0, \gamma > 0, \sigma > 0$ (which depend only on $\bar{m}, \bar{u}, \eta, f$ but not on the time horizon T), such that, if*

$$\delta < \sigma, \quad \|m_0(\cdot) - \bar{m}(\cdot)\|_{L^\infty(\mathbb{T}^n)} + \|u_T(\cdot) - \bar{u}(\cdot)\|_{W^{1,\infty}(\mathbb{T}^n)} < \gamma,$$

then there exists a (classical) solution (u^T, m^T) to (2.1) satisfying the following property:

$$\|m^T(t,\cdot) - \bar{m}(\cdot)\|_{L^\infty(\mathbb{T}^n)} \leq \frac{\bar{\varepsilon}}{2}\left(e^{-\sigma_1 t} + e^{-\sigma_2(T-t)}\right) \qquad \forall t \in [0,T],$$

where $\sigma_1 = \frac{\sigma-\delta}{n+1}$ and $\sigma_2 = \frac{\sigma+\delta}{n+1}$.

The existence of at least one trajectory that spends most of the time close to $(\bar{u}, \bar{m})$ is the best that one can expect in this framework; indeed, it is possible to construct examples of at least one solution sharing the same initial / terminal condition with the steady state, but deviating from it for most of the time horizon. In general, wild phenomena can occur without the presence of monotonicity: see for example [3].

2.4 Interpretation of the weak monotonicity assumption

In this section, we will provide an equivalent formulation for assumption (2.6). First, one can observe that it is equivalent to the validity of

$$\int_\Omega f_m(x,\bar{m})\mu^2 + \bar{m}\left|D\left(\frac{\mu}{\bar{m}}\right)\right|^2 + \delta\frac{\mu^2}{\bar{m}}\mathrm{d}x \geq \eta' \int_\Omega \frac{\mu^2}{\bar{m}}\mathrm{d}x, \tag{2.16}$$

for some $\eta' > 0$ and for all $\mu \in L^1(\Omega) \cap L^\infty(\Omega)$ such that $\int_\Omega \mu(x)\mathrm{d}x = 0$, $\mu/\bar{m} \in W^{1,2}(\bar{m}\mathrm{d}x)$. Assuming again $\delta = 0$ for simplicity, inequality (2.16) can be reformulated in terms of quotients of Rayleigh type, i.e. in terms of characterization of principal eigenvalues of some operators, as follows:

$$\inf_{\int_\Omega \mu = 0} \frac{\displaystyle\int_\Omega f_m(x,\bar{m})\mu^2 + \bar{m}\left|D\left(\frac{\mu}{\bar{m}}\right)\right|^2 \mathrm{d}x}{\displaystyle\int_\Omega \frac{\mu^2}{\bar{m}}\mathrm{d}x} \geq \eta' > 0. \tag{2.17}$$

In order to better interpret (2.17), let us introduce a new function

$$\psi = \frac{\mu}{\sqrt{\bar{m}}}.$$

This function is obviously orthogonal to $\sqrt{\bar{m}}$. If one rewrites (2.17) with respect to the new function above, we have

$$\inf_{\int_\Omega \sqrt{\bar{m}}\psi = 0} \frac{\langle L\psi, \psi\rangle_{L^2(\Omega)}}{\langle \psi, \psi\rangle_{L^2(\Omega)}} \geq \eta' > 0, \tag{2.18}$$

where

$$L(\psi) = f_m(x,\bar m)\bar m\psi - \frac{1}{\sqrt{\bar m}}\,\mathrm{div}\left(\bar m D\left(\frac{\psi}{\sqrt{\bar m}}\right)\right).$$

If such infimum is attained at some ψ, then (formally) the following equality holds

$$L(\psi) = \eta_1\psi + \ell\sqrt{\bar m}, \tag{2.19}$$

for some $\ell \in \mathbb{R}$. Via the change of variables $v = -\frac{\mu}{\bar m}$ and the identity $D\bar u = -\frac{D\bar m}{\bar m}$ one finds the PDE system solved by the triple (v,μ,ℓ) :

$$\begin{cases} -\Delta v + \langle Dv, D\bar u\rangle - f_m(x,\bar m)\mu = \eta_1 v + \ell \\ \Delta\mu + \mathrm{div}(\bar m Dv) + \mathrm{div}(\mu D\bar u) = 0, \\ \int_\Omega \mu = 0. \end{cases} \tag{2.20}$$

So, another way of interpreting (S) is the following:

the first (or principal) eigenvalue η_1, that is the smallest number η_1 such that (2.20) has nontrivial solutions, must be positive.

This interpretation can be exploited to produce non-trivial examples of stationary equilibria $(\bar u,\bar m)$ satisfying (2.16) or, equivalently, (2.6).

3 Perspectives

We now list some research lines that can be developed.

- Starting from the results in [4], it would be of some interest to address situations and models where the absence of monotonicity / convexity prevents from using the well-known tools from optimal control to get existence results. An interesting example is when f is a local anti-monotone function of the measure variable, like $f(x,m) = -C\log m$.
- Study another discounted problem, namely

$$\begin{cases} -\partial_t u - \Delta u + \frac{\delta|Du|^2}{2} - f(x,m) + \lambda u = 0, & \text{on } (0,T)\times\Omega \\ \partial_t m - \Delta m - \mathrm{div}(\delta m Du) = 0, & \text{on } (0,T)\times\Omega \\ m(0,x) = m_0(x), \qquad u(T,x) = u_T(x), & \text{on } \Omega, \end{cases} \tag{3.1}$$

  where $\delta,\lambda > 0$. This PDE system arises from the minimization problem

$$J(\alpha) = \mathbb{E}\int_0^T e^{-\lambda s}\left(\frac{1}{2\delta}|\alpha_s|^2 + f(X_s, m(s,X_s))\right)\mathrm{d}s + e^{-\lambda T}u_T(X_T),$$

  where each player controls his / her own state through

$$\mathrm{d}X_s = \alpha_s\mathrm{d}s + \sqrt{2}\mathrm{d}B_s \qquad \in \Omega.$$

Here one can notice the effect of two different discount factors, and the aim is to show that if δ and λ are large enough, the discount "convexifies" the coupling function f, allowing to compensate a possible lack of convexity in the original system and to have a full well-posedness theory for such PDE systems. This is an ongoing work with Prof. M. Cirant and Dr. E. Continelli.

References

[1] Bogachev, Vladimir I. and Krylov, Nicolai V. and Röckner, Michael and Shaposhnikov, Stanislav V., "Fokker-Planck-Kolmogorov equations". Mathematical Surveys and Monographs 207, American Mathematical Society, Providence, RI, 2015.

[2] P. Cardaliaguet, J.-M. Lasry, P.-L. Lions, and A. Porretta, *Long time average of mean field games*. Netw. Heterog. Media, 7(2): 279–301, 2012.

[3] A. Cesaroni and M. Cirant, *Brake orbits and heteroclinic connections for first order mean field games*. Trans. Amer. Math. Soc., 374(7): 5037–5070, 2021.

[4] M. Cirant and N. De Bernardi, *The local turnpike property in mean field control and games with quadratic hamiltonian*. 2026. arXiv:2511.18923.

[5] R. Dorfman, P. Samuelson, and R. Solow, "Linear programming and economic analysis". New York, McGraw-Hill, 1958.

[6] J.-M. Lasry and P.-L. Lions, *Jeux à champ moyen. II. Horizon fini et contrôle optimal*. C. R. Math. Acad. Sci. Paris, 343(10): 679–684, 2006.

[7] E. Trélat and E. Zuazua, *The turnpike property in finite-dimensional nonlinear optimal control*. J. Differential Equations, 258(1): 81–114, 2015.

[8] E. Trélat and E. Zuazua, *Turnpike in optimal control and beyond: a survey*. 2025. arXiv:2503.20342.

[9] J. von Neumann, *Über ein ökonomisches Gleichungssystem und eine Verallgemeinerung des brouwerschen Fixpunktsatzes*. K. Menger, ed., Ergebnisse eines Mathematischen Seminars, 1938.

An Introduction to Scattered Data Histopolation

Giacomo Cappellazzo (*)

Abstract. This survey is the summary of the talk proposed for the "Graduate Seminar" organized by the Doctoral School of Mathematical Sciences at the University of Padova. The full and detailed discussion is provided in "Scattered Data Histopolation in Averaging Kernel Hilbert Spaces" (L. Bruni Bruno, G. C., W. Erb, M. K. Esfahani).
Kernel-based methods offer a powerful and flexible mathematical framework for addressing histopolation problems. In histopolation, the available input data does not consist of pointwise function samples but of averages taken over intervals or higher-dimensional regions, and these mean values serve as a basis for reconstructing or approximating the target function. In the framework of kernel methods, we will introduce and study the so-called averaging kernel Hilbert spaces (AKHS's) for this purpose. Within this setting, we develop systematic construction principles for averaging kernels. Finally, we present numerical experiments that shed some light on the convergence behavior of the presented approach and demonstrate its practical effectiveness.

1 Introduction

Histopolation refers to approximation methods in which the mean values of functions over domains, such as intervals, cells, or higher-dimensional regions, are used as initial data rather than pointwise function samples used in interpolation. The term histopolation itself is a portmanteau formed from histogram and interpolation, and has been studied intensively in the context of splines, cf. (Siewer, 2008; Schoenberg, 1973; Subbotin, 1977) or (de Boor, 2001, Chapter VIII), in which spline approximants have to satisfy given area matching conditions. Using alternative basis functions, this idea was also studied for rational splines (Fischer & Oja, 2007), periodic splines (Delvos, 1987), or polynomials (Bruni Bruno et al., 2025; Bruni Bruno & Erb, 2024, 2025; Robidoux, 2006).

Also kernel-based methods have been used to solve histopolation-type problems. Compared to polynomials, kernel approximants turn out to be more stable and robust, and offer more flexibility when it comes to handle unstructured data sets. This was, for instance, observed for finite volume methods and in ENO/WENO schemes (Aboiyar et al.,

(*)Ph.D. course, Università di Padova, Dip. Matematica, via Trieste 63, I-35121 Padova, Italy. E-mail: cappella@math.unipd.it . Seminar held on 4 June 2026.

2010; Iske & Sonar, 1996; Sonar, 1996, 1997, 1998; Wendland, 2005). Another important field of application for kernel-based histopolation is the Radon transform, where the given initial data corresponds to integral values over entire lines or manifolds rather than intervals (De Marchi et al., 2016, 2018; Sironi, 2011; van den Boogaart et al., 2007). More generally, kernel-based histopolation arises as a special case in the context of generalized Hermite-Birkhoff interpolation, as for instance worked out in (Albrecht, 2024; Albrecht & Iske, 2025; Iske, 1995; Wu, 1992; Wendland, 2009) or (Wendland, 2004, Chapter 16).

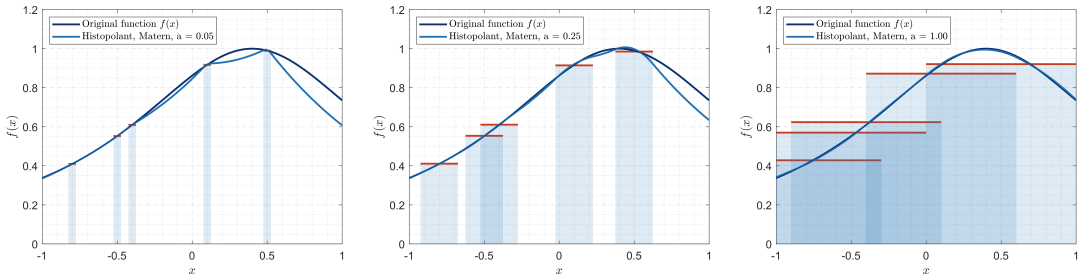


Figure 1: Histopolation of average values of the function $f(x) = \frac{1}{1+(x-0.4)^2}$ over 5 interval segments with length $a \in \{0.05, 0.25, 1\}$ using shifts of averaged Matérn functions as kernels, see Example 1 (i).

Common to all kernel-based methods is the use of an underlying reproducing kernel Hilbert space (RKHS), sometimes also referred to as native space, which serves as the fundamental approximation space for formulating (generalized) interpolation problems. These spaces contain a reproducing kernel that enables to represent point-evaluation functionals as elements of the Hilbert space itself.

In histopolation, point evaluations are not essential. Instead, it suffices that averages over the considered subdomains are continuous functionals in the underlying approximation space, a substantially weaker assumption. To obtain a theoretical foundation for histopolation, we will therefore introduce and systematically study averaging kernel Hilbert spaces (AKHS's), in which only the specified averaging functionals are required to be continuous. This new framework enables the approximation of more general, also non-continuous functions. A representative example is the space $L_2(\mathbb{R}^d)$ of square-integrable functions. Although it is not a RKHS, it can be interpreted as an AKHS. In this sense, AKHS's extend naturally the classical native-space theory and allow to leave the pure RKHS framework while still remaining within a Hilbert space setting.

Nevertheless, AKHS's and RKHS's are closely related. This relation between an AKHS and its associated RKHS has deep theoretical and practical implications for histopolation: it provides a way to characterize averaging kernels through known descriptions of reproducing kernels, it allows histopolation matrices to be stated explicitly in terms of the reproducing kernel, and it yields criteria for unisolvence based on the corresponding criteria for the reproducing kernel (Section 4).

In Section 5 we obtain convergence criteria for the mean values of the histopolants, again as a consequence of the isomorphism between the AKHS and its associated RKHS.

In the final Section 6 we show how these methods can be applied to image processing tasks, such as image compression through pixel binning and image upscaling.

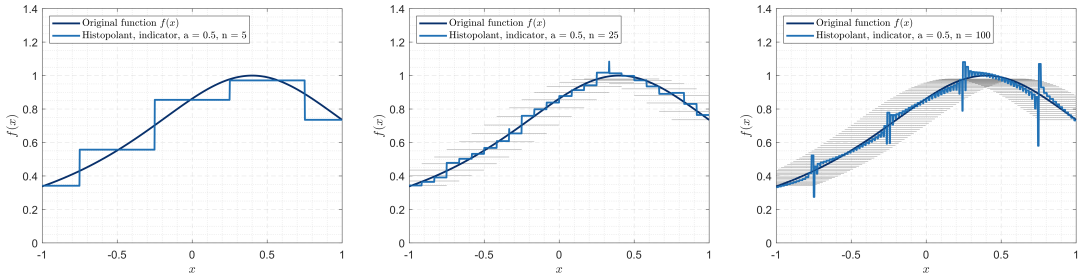


Figure 2: Histopolation of average values of the function $f(x) = \frac{1}{1+(x-0.4)^2}$ over $n \in \{5, 25, 100\}$ uniform intervals of length $a = 0.5$ using the indicator functions in (4) as averaging kernels.

2 Averaging Kernel Hilbert Spaces

In the following, we assume that Ω is a subset of $\mathbb{R}^d$ and $\mathcal{H}$ is a Hilbert space of real-valued functions on Ω, endowed with pointwise addition and scalar multiplication of the functions. The relative inner product is denoted by $\langle \cdot, \cdot \rangle_{\mathcal{H}}$.

We consider a collection $\{\omega_\tau \ : \ \tau \in \mathcal{T}\}$ of Lebesgue measurable subsets in Ω, where $\mathcal{T}$ is an arbitrary set of indices. For the Hilbert space $\mathcal{H}$, we then consider the linear *averaging functionals*

$$\lambda_\tau : f \mapsto \frac{1}{|\omega_\tau|} \int_{\omega_\tau} f(x) \mathrm{d}x, \quad \text{for all } f \in \mathcal{H}, \ \tau \in \mathcal{T},$$

where $|\omega_\tau|$ denotes the Lebesgue measure of the set ω_τ, and assume that they are continuous on $\mathcal{H}$. These functionals provide the averages of a function f over the subsets ω_τ. As the functionals λ_τ are continuous, the Riesz representation theorem ensures that we can represent them as dual elements $A(\cdot, \tau) \in \mathcal{H}$ in the Hilbert space such that

$$\lambda_\tau(f) = \langle f, A(\cdot, \tau) \rangle_{\mathcal{H}}.$$

We refer to the kernel function $A : \Omega \times \mathcal{T} \to \mathbb{R}$ as *averaging kernel*. Further, we call the subspace

$$\mathcal{N}_A(\Omega) := \overline{\operatorname{span}\{A(\cdot, \tau) \ : \ \tau \in \mathcal{T}\}} \subseteq \mathcal{H}$$

an *averaging kernel Hilbert space (AKHS)*. In particular, in this Hilbert space the following averaging property holds true:

$$\frac{1}{|\omega_\tau|} \int_{\omega_\tau} f(x) \mathrm{d}x = \lambda_\tau(f) = \langle f, A(\cdot, \tau) \rangle_{\mathcal{H}} \quad \text{for all } f \in \mathcal{N}_A(\Omega).$$

If we regard $A(\cdot, \tau)$ as a function in $\mathcal{N}_A(\Omega)$, we can as well calculate its averages and get

$$\frac{1}{|\omega_\rho|} \int_{\omega_\rho} A(y, \tau) \mathrm{d}y = \lambda_\rho(A(\cdot, \tau)) = \langle A(\cdot, \tau), A(\cdot, \rho) \rangle_{\mathcal{H}},$$

where $A(\cdot, \rho) \in \mathcal{N}_A(\Omega)$ is the element in $\mathcal{N}_A(\Omega)$ associated with the averaging functional λ_ρ. This allows us to define a new associated kernel $K : \mathcal{T} \times \mathcal{T} \to \mathbb{R}$ by

$$K(\tau, \rho) := \langle A(\cdot, \tau), A(\cdot, \rho) \rangle_{\mathcal{H}}.$$

Based on this definition, we can see that $K : \mathcal{T} \times \mathcal{T} \to \mathbb{R}$ is a symmetric and positive semi-definite kernel on the index set $\mathcal{T}$. Specifically, for every $n \in \mathbb{N}$, $\tau_1, \ldots, \tau_n \in \mathcal{T}$, and $c_1, \ldots, c_n \in \mathbb{R}$, we have:

$$\sum_{i,j=1}^{n} c_i c_j K(\tau_i, \tau_j) = \left\| \sum_{i=1}^{n} c_i A(\cdot, \tau_i) \right\|_{\mathcal{H}}^2 \geq 0.$$

Further, if the functions $A(\cdot, \tau)$ are linearly independent, the kernel K is positive definite, implying a strict inequality in the latter formula if the coefficients c_i are not vanishing simultaneously. We can associate to the kernel K a reproducing kernel Hilbert space $\mathcal{N}_K(\mathcal{T})$ that is given as the completion

$$\mathcal{N}_K(\mathcal{T}) := \overline{\operatorname{span}\{K(\cdot, \tau) \ : \ \tau \in \mathcal{T}\}},$$

in which the inner product of two pre-Hilbert space elements written as $f(\tau) = \sum_{i=1}^{n} c_i K(\tau, \tau_i)$ and $g(\tau) = \sum_{i=1}^{n} d_i K(\tau, \tau_i)$ is defined as

$$\langle f, g \rangle_{\mathcal{N}_K(\mathcal{T})} := \sum_{i,j=1}^{n} c_i d_j K(\tau_i, \tau_j).$$

We then obtain the following isometry, which is analogous to those established in the context of Hermite-Birkhoff interpolation inside RKHS's (see Wendland, 2004, Theorem 16.8). Novel in our setting is that within the framework of AKHS's we are no longer restricted to start in an initial RKHS.

Theorem 1 *Assume that the functionals $\{\lambda_\tau \ : \ \tau \in \mathcal{T}\}$ are continuous in the space $\mathcal{H}$. Then, the linear operator $\Lambda : \mathcal{N}_A(\Omega) \to \mathcal{N}_K(\mathcal{T})$ given as $(\Lambda f)(\tau) = \lambda_\tau(f)$ provides an isometric isomorphism between the averaging kernel Hilbert space $\mathcal{N}_A(\Omega)$ and the reproducing kernel Hilbert space $\mathcal{N}_K(\mathcal{T})$. In particular, for the respective inner products we have*

$$\langle f, g \rangle_{\mathcal{H}} = \langle \Lambda f, \Lambda g \rangle_{\mathcal{N}_K(\mathcal{T})}.$$

3 Construction and examples of AKHS

In this section we provide several explicit examples of averaging kernel Hilbert spaces. In particular, we show that spaces which do not admit a reproducing kernel may nevertheless admit an averaging kernel. A particular focus is given to the construction of AKHS's from given RKHS's and on uniform AKHS's in which the averaging domains are shifts of a single domain.

3.1 Square integrable functions

As a first example, we consider the Hilbert space $\mathcal{H} = L_2(\Omega)$ equipped with the usual inner product $\langle f, g \rangle_{\mathcal{H}} = \int_\Omega f(x) g(x) \mathrm{d}x$. This Hilbert space does not have a reproducing

kernel, but we can build averaging kernels for $L_2(\Omega)$. For a given set $\{\omega_\tau \ : \ \tau \in \mathcal{T}\}$ of measureable subdomains in Ω the indicator functions

$$A(x,\tau) = \frac{1}{|\omega_\tau|}\chi_{\omega_\tau}(x),$$

define an averaging kernel and a respective AKHS inside $L_2(\Omega)$. The associated reproducing kernel K on $\mathcal{T}\times\mathcal{T}$ is given as

$$K(\rho,\tau) = \frac{1}{|\omega_\tau||\omega_\rho|}|\omega_\rho \cap \omega_\tau|.$$

3.2 Averaging kernels generated in reproducing kernel Hilbert spaces

As an important construction strategy for averaging kernels, we consider as initial Hilbert space $\mathcal{H}$ a RKHS $\mathcal{N}_\Phi(\Omega)$ generated by a symmetric positive definite reproducing kernel $\Phi \in C(\Omega\times\Omega)$ on a domain $\Omega \subseteq \mathbb{R}^d$. This strategy will allow us to obtain the averaging kernels for a large class of Hilbert spaces. Since RKHS's are typically linked to smoother function spaces, also the AKHS's build on these spaces will have a respectively larger smoothness.

Consider a collection $\{\omega_\tau \ : \ \tau \in \mathcal{T}\}$ of subdomains in Ω. As we assume that the reproducing kernel Φ is continuous, and since the point evaluation functionals in $\mathcal{N}_\Phi(\Omega)$ are continuous, by the positive definiteness of Φ we immediately get, for any $f \in \mathcal{N}_\Phi(\Omega)$, the bound

$$\left|\frac{1}{|\omega_\tau|}\int_{\omega_\tau} f(x)\mathrm{d}x\right| \leq \frac{1}{|\omega_\tau|}\int_{\omega_\tau}|\langle f,\Phi(\cdot,x)\rangle_{\mathcal{N}_\Phi(\Omega)}|\mathrm{d}x \leq \frac{\|f\|_{\mathcal{N}_\Phi(\Omega)}}{|\omega_\tau|}\int_{\omega_\tau}\sqrt{\Phi(x,x)}\mathrm{d}x \leq \|\Phi\|_\infty^{\frac{1}{2}}\|f\|_{\mathcal{N}_\Phi(\Omega)}.$$

This implies that all averaging functionals λ_τ in $\mathcal{N}_\Phi(\Omega)$ are well-defined and continuous. We can therefore introduce an averaging kernel $A(x,\tau)$, and get

$$\langle f, A(\cdot,\tau)\rangle_{\mathcal{N}_\Phi(\Omega)} = \frac{1}{|\omega_\tau|}\int_{\omega_\tau} f(x)\mathrm{d}x = \frac{1}{|\omega_\tau|}\int_{\omega_\tau}\langle f,\Phi(\cdot,x)\rangle_{\mathcal{N}_\Phi(\Omega)}\mathrm{d}x = \left\langle f, \frac{1}{|\omega_\tau|}\int_{\omega_\tau}\Phi(\cdot,x)\mathrm{d}x\right\rangle_{\mathcal{N}_\Phi(\Omega)}$$

for all $f \in \mathcal{N}_\Phi(\Omega)$. Therefore, the averaging kernel has the form

$$A(x,\tau) = \frac{1}{|\omega_\tau|}\int_{\omega_\tau}\Phi(x,y)\mathrm{d}y, \tag{1}$$

and it can be calculated in terms of mean values of the reproducing kernel Φ. Furthermore, as a consequence of Fubini's theorem, the reproducing kernel $K(\rho,\tau)$ can also be formulated as

$$K(\rho,\tau) = \frac{1}{|\omega_\rho|}\int_{\omega_\rho}A(x,\tau)\mathrm{d}x = \frac{1}{|\omega_\rho|}\frac{1}{|\omega_\tau|}\int_{\omega_\rho}\int_{\omega_\tau}\Phi(x,y)\,\mathrm{d}y\mathrm{d}x. \tag{2}$$

3.3 Uniform averaging kernels based on translates of a single domain

As a further more structured example of an AKHS, we consider the spaces $L_2(\mathbb{R}^d)$ with the averaging functionals given by the translates of a single subset $\omega \subseteq \mathbb{R}^d$. For further simplicity, we suppose that ω is a symmetric set centered at the origin, i.e., with $x \in \omega$ also $-x \in \omega$. Indexed by the set $\mathcal{T} = \mathbb{R}^d$, we then consider the subdomains $\{\omega_y = y + \omega \ : y \in \mathbb{R}^d\}$. Based on our assumptions, the averaging operations can now be formulated as

$$\lambda_y(f) = \frac{1}{|\omega|} \int_{y+\omega} f(x)\mathrm{d}x = \left(f * \frac{1}{|\omega|}\chi_\omega \right)(y). \tag{3}$$

In particular, the averaging functionals correspond to the point evaluation of the convolution of the function f with the normalized indicator function $\frac{1}{|\omega|}\chi_\omega$. Several averaging kernels can be constructed explicitly in this framework.

3.3.1 B-splines in the univariate case

As a first simple example in the univariate case, we start with the Hilbert space $L_2(\mathbb{R})$ and the domain $\omega = [-a/2, a/2]$ with length $a > 0$. To obtain the averaging operation in (3), the corresponding averaging kernel in $L_2(\mathbb{R})$ is given by the indicator function

$$A_1(x,y) = \frac{1}{a}\chi_{[y-a/2,y+a/2]}(x) = \begin{cases} 1/a & \text{if } |x-y| \leq a/2 \\ 0 & \text{if } |x-y| > a/2 \end{cases}. \tag{4}$$

Moreover, as respective associated reproducing kernel we get

$$K_1(x,y) = \frac{1}{a}\left(1 - \frac{|x-y|}{a}\right)_+,$$

where $(x-y)_+ = \max(0, x-y)$ is known as positive part. The kernels $A_1(x,y)$ and $K_1(x,y)$ can be written in terms of the central B-splines $M_1(x)$ and $M_2(x)$ as

$$A_1(x,y) = \frac{1}{a}M_1\left(\frac{x-y}{a}\right), \quad K_1(x,y) = \frac{1}{a}M_2\left(\frac{x-y}{a}\right).$$

We can extend this using the n-th central B-spline $M_n(x)$. It is generally defined as the $(n-1)$-fold convolution of the indicator function $M_1(x) = \chi_{[-1/2,1/2]}(x)$, such that

$$M_n(x) = \underbrace{(M_1 * M_1 * \cdots * M_1)}_{(n-1 \text{ convolutions})}(x).$$

By iteratively applying the convolution with the indicator function $\frac{1}{|\omega|}\chi_\omega$ to the kernel $K_1(x,y)$, we can generate further AKHS's with the averaging kernels

$$A_n(x,y) = \frac{1}{a}M_{2n-1}\left(\frac{x-y}{a}\right), \quad n \in \mathbb{N}.$$

The corresponding associate reproducing kernels $K_n(x,y)$ are then given as

$$K_n(x,y) = \frac{1}{a}M_{2n}\left(\frac{x-y}{a}\right), \quad n \in \mathbb{N}.$$

3.3.2 Radial uniform averaging kernels for $d = 1$

A uniform averaging kernel $A : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is called *radial* if

$$A(x, y) = \alpha(\|x - y\|_2)$$

for some univariate function $\alpha : [0, \infty) \to \mathbb{R}$. For a uniform AKHS with a radial kernel $A(x, y)$ and a ball ω as an underlying averaging domain, also the associated reproducing kernel $K(x, y)$ is radial and can be written as

$$K(x, y) = \kappa(\|x - y\|_2)$$

with a univariate function $\kappa : [0, \infty) \to \mathbb{R}$.

In the univariate setting $d = 1$, and using the domain $\omega = [-a/2, a/2]$ with interval length a, the respective averaging kernel can be written as $A(x, y) = \alpha(x-y)$ by extending α evenly to the entire real axis. The same holds true for the reproducing kernel $K(x, y) = \kappa(x - y)$, that can now be expressed in terms of the even function

$$\kappa(x) = \left(\alpha * \frac{1}{a}\chi_{[-a/2,a/2]}\right)(x).$$

As the reproducing kernel K is positive semi-definite, the function κ itself is a positive semi-definite function. If the underlying space $\mathcal{H}$ is a native space $\mathcal{N}_\Phi(\Omega)$ based on a univariate radial kernel Φ, such that

$$\Phi(x, y) = \phi(x - y),$$

the uniform averaging kernel A and the reproducing kernel K can be calculated in terms of ϕ as

$$\alpha(x) = \phi * \frac{1}{a}\chi_{[-a/2,a/2]}(x), \quad \kappa(x) = \phi * \frac{1}{a}\chi_{[-a/2,a/2]} * \frac{1}{a}\chi_{[-a/2,a/2]}(x).$$

In particular, if the function ϕ is positive definite, also κ is positive definite. If the first two anti-derivatives of ϕ exist, the functions α and κ can be stated explicitly by the fundamental theorem of calculus.

Lemma 1 *Let $I_\phi(x)$ and $I^2_\phi(x)$ denote the first and second anti-derivative of the univariate function ϕ that generates the native space $\mathcal{N}_\Phi(\Omega)$. Then, the radial functions α and κ of the averaging kernel A and the reproducing kernel K can be written as*

$$\alpha(x) = \frac{1}{a}\left(I_\phi(x + a/2) - I_\phi(x - a/2)\right), \quad \kappa(x) = \frac{1}{a^2}\left(I^2_\phi(x + a) + I^2_\phi(x - a) - 2I^2_\phi(x)\right),$$

i.e., α and κ are first- and second-order symmetric finite differences of the first and second anti-derivative of ϕ with respect to the step size a, respectively.

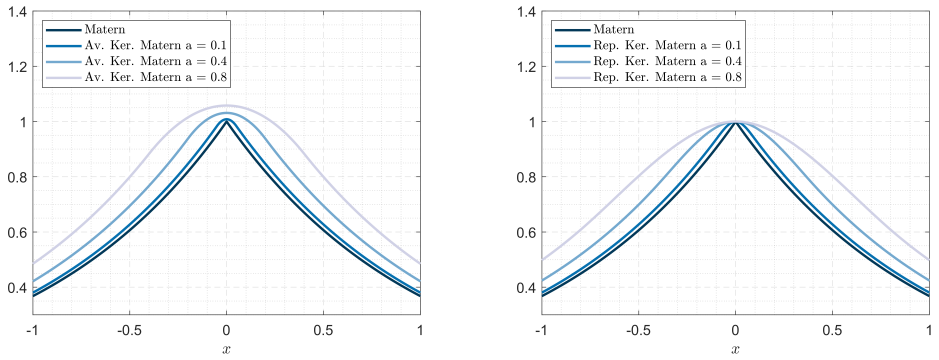


Figure 3: The averaging and reproducing kernel based on the Matérn function. Left: the Matérn function $\phi(x) = e^{-\lambda|x|}$ compared to the averaging kernels $\alpha(x)/\kappa(0)$ given in Example 1 (i) for three interval lengths $a \in \{0.1, 0.4, 0.8\}$. Right: the respective normalized reproducing kernels $\kappa(x)/\kappa(0)$.

Lemma 1 gives a convenient way to compute averaging kernels explicitly, as the next example shows.

Example 1

(i) For the Matérn function $\phi(x) = e^{-\lambda|x|}$, we get

$$\alpha(x) = \begin{cases} \frac{1}{\lambda a}(2 - e^{\lambda(|x|-a/2)} - e^{-\lambda(|x|+a/2)}) & \text{if } |x| \leq a/2, \\ \frac{1}{\lambda a}(e^{-\lambda(|x|-a/2)} - e^{-\lambda(|x|+a/2)}) & \text{if } |x| > a/2, \end{cases}$$

and

$$\kappa(x) = \begin{cases} \frac{1}{\lambda^2 a^2}(2\lambda(a - |x|) + e^{\lambda(|x|-a)} + e^{-\lambda(|x|+a)} - 2e^{-\lambda(|x|)}) & \text{if } |x| \leq a, \\ \frac{1}{\lambda^2 a^2}(e^{-\lambda(|x|-a)} + e^{-\lambda(|x|+a)} - 2e^{-\lambda(|x|)}) & \text{if } |x| > a. \end{cases}$$

(ii) Starting with the anti-derivatives I_ϕ^2 and I_ϕ, Lemma 1 can be applied also in a slightly different way to obtain averaging and reproducing kernels explicitly. As an example, it is well-known that $I_\phi^2(x) = -\frac{1}{2\lambda}e^{-\lambda x^2}$ with $\lambda > 0$ is the first and second anti-derivative of the functions

$$I_\phi(x) = xe^{-\lambda x^2}, \quad \text{and} \quad \phi(x) = \left(1 - 2\lambda x^2\right) e^{-\lambda x^2}.$$

The function $\phi(x)$ is positive definite and known as Mexican hat wavelet. Using Lemma 1, we immediately get the averaging kernel for ϕ as

$$\alpha(x) = \left(\frac{1}{2} + \frac{x}{a}\right) e^{-\lambda(x+\frac{a}{2})^2} + \left(\frac{1}{2} - \frac{x}{a}\right) e^{-\lambda(x-\frac{a}{2})^2}$$

and the respective reproducing kernel as

$$\kappa(x) = \frac{1}{2\lambda a^2}\left(2e^{-\lambda x^2} - e^{-\lambda(x+a)^2} - e^{-\lambda(x-a)^2}\right).$$

4 Solution of histopolation problems

The AKHS setting turns out to be the right theoretical framework to solve histopolation problems with a kernel method. To state and solve the problem, we assume that $\mathcal{N}_A(\Omega) \subseteq \mathcal{H}$ is an averaging kernel Hilbert space based on averages created on the subdomains $\{\omega_\tau \ : \ \tau \in \mathcal{T}\}$. We further assume that we are given the mean values

$$\lambda_{\tau_i}(f) = \frac{1}{|\omega_{\tau_i}|} \int_{\omega_{\tau_i}} f(x)\mathrm{d}x \tag{5}$$

of the function $f \in \mathcal{N}_A(\Omega)$ on n subdomains $\{\omega_{\tau_1}, \ldots, \omega_{\tau_n}\}$ based on the indices $\mathcal{T}_n = \{\tau_1, \ldots, \tau_n\} \subseteq \mathcal{T}$. Our aimed-at approximant s_f of the function f should retain these mean values in the histopolation space

$$\mathcal{N}_A(\Omega, \mathcal{T}_n) = \operatorname{span}\{A(\cdot, \tau_j) \ ; \ j \in \{1, \ldots n\}\} \subseteq \mathcal{N}_A(\Omega).$$

More precisely, we want the approximant

$$s_f(x) = \sum_{j=1}^{n} c_j A(x, \tau_j) \in \mathcal{N}_A(\Omega, \mathcal{T}_n) \tag{6}$$

to satisfy the n histopolation conditions

$$\lambda_{\tau_i}(s_f) = \frac{1}{|\omega_{\tau_i}|} \int_{\omega_{\tau_i}} s_f(x)\mathrm{d}x = \frac{1}{|\omega_{\tau_i}|} \int_{\omega_{\tau_i}} f(x)\mathrm{d}x = \lambda_{\tau_i}(f), \qquad i \in \{1, \ldots n\}.$$

These n conditions can be formulated in terms of a linear system of equations

$$\mathbf{K}\,\mathbf{c} = \boldsymbol{\lambda}, \tag{7}$$

where the vector $\mathbf{c} = [c_1 \cdots c_n]^\intercal$ contains the expansion coefficients of the histopolant s_f,

$$\boldsymbol{\lambda} = [\lambda_{\tau_1}(f) \ \cdots \ \lambda_{\tau_n}(f)]^\intercal$$

contains the histopolation data(5) and $\mathbf{K}$ is the $n \times n$ positive semi-definite matrix with the entries

$$\mathbf{K}_{i,j} = \frac{1}{|\omega_{\tau_i}|} \int_{\omega_{\tau_i}} A(x, \tau_j)dx = K(\tau_i, \tau_j). \tag{8}$$

In an AKHS, the linear independence of the functionals $\{\lambda_{\tau_1}, \ldots, \lambda_{\tau_n}\}$ is equivalent to the linear independence of the elements $\{A(\cdot, \tau_1), \ldots, A(\cdot, \tau_n)\}$. The averaging property of the AKHS further implies that

$$\left\| \sum_{i=1}^{n} \alpha_i A(\cdot, \tau_i) \right\|_{\mathcal{H}}^2 = \left\langle \sum_{i=1}^{n} \alpha_i A(\cdot, \tau_i), \sum_{j=1}^{m} \alpha_j A(\cdot, \rho_j) \right\rangle_{\mathcal{H}} = \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \alpha_j K(\tau_i, \rho_j),$$

implying that the matrix $\mathbf{K}$ is positive definite and, thus, invertible if the functionals $\{\lambda_{\tau_1}, \ldots, \lambda_{\tau_n}\}$ are linearly independent. The developed theory on averaging kernel Hilbert spaces implies now the following result.

Theorem 2 *If $\mathcal{N}_A(\Omega)$ is an averaging kernel Hilbert space with linearly independent averaging functionals $\{\lambda_{\tau_1}, \ldots, \lambda_{\tau_n}\}$, then the matrix $\mathbf{K}$ is positive definite, and the system (7) has a unique solution. Further, the histopolant $s_f \in \mathcal{N}_A(\mathcal{T}_n)$ in (6) can be calculated explicitly as*

$$s_f(x) = \mathbf{a}(x)^\intercal \mathbf{K}^{-1} \boldsymbol{\lambda},$$

where $\mathbf{a}(x)$ denotes the vector

$$\mathbf{a}(x) = [A(x, \tau_1) \ \cdots \ A(x, \tau_n)]^\intercal.$$

Remark 1 The combination of Theorem 2 with Theorem 1 reveals the following fundamental connection between interpolation and histopolation problems: solving a histopolation problem in an AKHS $\mathcal{N}_A(\Omega)$ is equivalent to solving a respective interpolation problem in the RKHS $\mathcal{N}_K(\mathcal{T})$.

The linear independence of the averaging functionals $\{\lambda_{\tau_1}, \ldots, \lambda_{\tau_n}\}$ is a fundamental prerequisite for Theorem 2. Since these functionals essentially depend on the domains $\{\omega_{\tau_1}, \ldots, \omega_{\tau_n}\}$, we aim to identify conditions on these domains that guarantee the linear independence of the kernel functions $\{A(\cdot, \tau_1), \ldots, A(\cdot, \tau_n)\}$. To this end, following Section 3.2, we assume now that the averaging kernel A is constructed upon a continuous symmetric positive definite kernel Φ. For such a kernel we can introduce the compact integral operator $T : L_2(\Omega) \to \mathcal{N}_\Phi(\Omega) \subseteq L_2(\Omega)$, defined by

$$T(v)(x) = \int_\Omega \Phi(x, y) v(y) \mathrm{d}y, \qquad v \in L_2(\Omega), \quad x \in \Omega.$$

If $\Omega \subset \mathbb{R}^d$ is compact, the integral operator T maps $L_2(\Omega)$ continuously into the native space $\mathcal{N}_\Phi(\Omega)$. It is the adjoint of the embedding operator of the RKHS $\mathcal{N}_\Phi(\Omega)$ into $L_2(\Omega)$, i.e., it satisfies

$$\langle f, v \rangle_{L_2(\Omega)} = \langle f, T(v) \rangle_{\mathcal{N}_\Phi(\Omega)}. \tag{9}$$

Furthermore, the range of T is dense in $\mathcal{N}_\Phi(\Omega)$, as shown in (Wendland, 2004, Proposition 10.28).

If the averaging kernel A is constructed upon the positive definite kernel Φ, the corresponding reproducing kernel K is determined by (2), and the entries of the histopolation matrix $\mathbf{K}$ in (8) based on the finite collection of subsets $\{\omega_{\tau_1}, \ldots, \omega_{\tau_n}\}$ can be explicitly written as

$$\mathbf{K}_{i,j} = \frac{1}{|\omega_{\tau_i}|} \frac{1}{|\omega_{\tau_j}|} \int_{\omega_{\tau_i}} \int_{\omega_{\tau_j}} \Phi(x, y) \mathrm{d}x \mathrm{d}y. \tag{10}$$

We can now show the following result, which relates the unisolvence of the histopolation problem with the linear independence of the characteristic functions of the averaging domains in the space $L_2(\Omega)$.

Theorem 3 *Let $\Phi \in C(\Omega \times \Omega)$ be a symmetric positive definite kernel on the compact set Ω. We assume further that the RKHS $\mathcal{N}_\Phi(\Omega)$ is dense in $L_2(\Omega)$ and that the characteristic functions $\{\chi_{\omega_{\tau_1}}, \dots, \chi_{\omega_{\tau_n}}\}$ are linearly independent in $L_2(\Omega)$. Then, the histopolation matrix $\mathbf{K}$ in (10) is symmetric and positive definite, and, in particular, invertible.*

Proof. The symmetry of $\mathbf{K}$ follows from the symmetry of the reproducing kernel K, can however be also derived directly from (10) by using the symmetry of Φ and Fubini's theorem. To prove that the matrix $\mathbf{K}$ is positive definite, it is convenient to write the corresponding entries in terms of the characteristic functions $\chi_{\omega_{\tau_i}}$ and the integral operator T:

$$\begin{aligned}|\omega_{\tau_i}||\omega_{\tau_j}|\mathbf{K}_{i,j} &= \int_{\omega_{\tau_i}} \int_{\omega_{\tau_j}} \Phi(x,y)\mathrm{d}x\mathrm{d}y = \\ &= \int_\Omega \left(\chi_{\omega_{\tau_i}}(y) \int_\Omega \Phi(x,y)\chi_{\omega_{\tau_j}}(x)\mathrm{d}x \right) \mathrm{d}y = \langle \chi_{\omega_{\tau_i}}, T(\chi_{\omega_{\tau_j}}) \rangle_{L_2(\Omega)}.\end{aligned}$$

We consider now $\alpha \in \mathbb{R}^n \backslash \{0\}$ and show that $\mathbf{K}$ is positive definite, i.e., that $\sum_{i,j=1}^n \alpha_i \alpha_j \mathbf{K}_{i,j} > 0$. Using the identity above, we get

$$\begin{aligned}\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \mathbf{K}_{i,j} &= \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \left\langle \frac{1}{|\omega_{\tau_i}|}\chi_{\omega_{\tau_i}}, T\left(\frac{1}{|\omega_{\tau_j}|}\chi_{\omega_{\tau_j}} \right) \right\rangle_{L_2(\Omega)} = \\ &= \sum_{i=1}^n \alpha_i \left\langle \frac{1}{|\omega_{\tau_i}|}\chi_{\omega_{\tau_i}}, T\left(\sum_{j=1}^n \frac{\alpha_j}{|\omega_{\tau_j}|}\chi_{\omega_{\tau_j}} \right) \right\rangle_{L_2(\Omega)} = \\ &= \left\langle \sum_{i=1}^n \frac{\alpha_i}{|\omega_{\tau_i}|}\chi_{\omega_{\tau_i}}, T\left(\sum_{j=1}^n \frac{\alpha_j}{|\omega_{\tau_j}|}\chi_{\omega_{\tau_j}} \right) \right\rangle_{L_2(\Omega)} = \langle v, T(v) \rangle_{L_2(\Omega)},\end{aligned}$$

where $v = \sum_{i=1}^n \frac{\alpha_i}{|\omega_{\tau_i}|}\chi_{\omega_{\tau_i}}$ and $v \neq 0$ because $\{\omega_{\tau_1}, \dots, \omega_{\tau_n}\}$ are linearly independent. By the identity (9) shown in (Wendland, 2004, Proposition 10.28), we therefore get

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \mathbf{K}_{i,j} = \langle v, T(v) \rangle_{L_2(\Omega)} = \langle T(v), T(v) \rangle_{\mathcal{N}_\Phi(\Omega)} = \|T(v)\|^2_{\mathcal{N}_\Phi(\Omega)} > 0,$$

where we have strict positivity due to the injectivity of the integral operator T. The injectivity of T follows from the fact that the range of the adjoint operator (the embedding operator from $\mathcal{N}_\Phi(\Omega)$ into $L_2(\Omega)$) corresponds to $\mathcal{N}_\Phi(\Omega)$ and the assumption that $\mathcal{N}_\Phi(\Omega)$ is dense in $L_2(\Omega)$. □

Theorem 3 is quite valuable for obtaining theoretical guarantees of the invertibility of the histopolation matrix $\mathbf{K}$. In fact, for many well-known positive definite kernels, such as

the Matérn and Wendland kernels (cf. Wendland, 2004), it is known that the associated RKHS's are equivalent to Sobolev spaces. Therefore, their restrictions to a domain Ω are dense in $L_2(\Omega)$. In such cases it is then sufficient to verify that the characteristic functions $\{\chi_{\omega_{\tau_1}}, \dots, \chi_{\omega_{\tau_n}}\}$ are linearly independent in $L_2(\Omega)$.

In the following, we establish a sufficient condition for the linear independence of the characteristic functions $\{\chi_{\omega_{\tau_1}}, \dots, \chi_{\omega_{\tau_n}}\}$ in $L_2(\Omega)$. This result can be seen as a minor extension of the criterion presented in (Albrecht, 2024, Theorem 4.7), where linear independence was considered in the space of compactly supported distributions.

Theorem 4 *Let $\{\omega_{\tau_1}, \dots, \omega_{\tau_n}\}$ be a collection of bounded domains in $\Omega \subseteq \mathbb{R}^d$. Assume that they are ordered in such a way that there exist open balls $o_\varepsilon^{(i)}$, $i \in \{1, \dots, n\}$, with radius $\varepsilon > 0$ such that*

$$o_\varepsilon^{(i)} \subseteq \omega_{\tau_i}, \quad o_\varepsilon^{(i)} \cap \bigcup_{j=i+1}^{n} \omega_{\tau_j} = \emptyset, \quad i \in \{1, \dots, n\}. \tag{11}$$

Then, the characteristic functions $\{\chi_{\omega_{\tau_1}}, \dots, \chi_{\omega_{\tau_n}}\}$ are linearly independent in $L_2(\Omega)$.

Example 2 On a compact interval $I \subseteq \mathbb{R}$, consider the subintervals $\{\omega_{y_1}, \dots, \omega_{y_n}\}$ given as $\omega_{y_i} = [-a/2, a/2] + y_i$ with ordered centers $y_1 < y_2 < \cdots < y_n$ and uniform interval length $a > 0$. Then, by Theorem 4, the characteristic functions $\{\chi_{\omega_{y_1}}, \dots, \chi_{\omega_{y_n}}\}$ are linearly independent in $L_2(I)$.

(i) The Matérn kernel function $\phi(x) = e^{-\lambda|x|}$ considered in Example 1 (i) generates a RKHS which is norm-equivalent to the Sobolev space $H^1(I)$. Therefore, $\mathcal{N}_\phi(I)$ is dense in $L_2(I)$. For the segments introduced above, we can therefore apply Theorem 3 and obtain that the respective histopolation matrix $\mathbf{K}$ is invertible. This guarantees the unisolvence of the histopolation problem in the example illustrated in Fig. 1.

(ii) For the Hilbert space $L_2(I)$ we know directly that the indicator functions $\{\chi_{\omega_{y_1}}, \dots, \chi_{\omega_{y_n}}\}$ are linearly independent. For this, also the averaging kernel $A_1(x, y) = \frac{1}{a}\chi_{\omega_y}(x)$ introduced in (4) provides linearly independent basis functions $\{A_1(x, y_1), \dots A_1(x, y_n)\}$ and the corresponding histopolation matrix $\mathbf{K}$ is invertible. This ensures the unisolvence of the histopolation problem shown in Fig. 2.

5 Uniform error estimates for average values

For histopolation, a natural type of convergence is the uniform convergence of the mean values. In fact, the isometry of Theorem 1 will guarantee that for a large class of functions. The average values $\lambda_\tau(s_f)$ converge towards the values $\lambda_\tau(f)$ if the domains ω_τ are sufficiently dense.

To obtain respective error bounds, we introduce the Lagrange basis functions $\ell_j(x)$, $j \in \{1, \dots, n\}$ as the elements in the space $\mathcal{N}_A(\Omega, \mathcal{T}_n)$ that satisfy the cardinal histopolation conditions

$$\lambda_{\tau_i}(\ell_j) = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

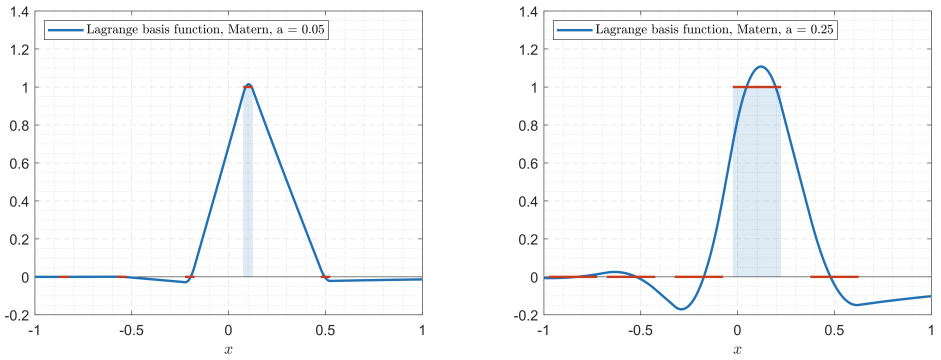


Figure 4: Lagrange basis functions for histopolation on 5 segments (red) with length $a = 0.05$ (left) and $a = 0.25$ (right) using the one-dimensional averaged Matérn function in Example 1 (i).

If the conditions of Theorem 2 are satisfied, the functions $\ell_j(x)$, $j \in \{1, \dots, n\}$, are uniquely determined and form a basis of the n-dimensional space $\mathcal{N}_A(\Omega, \mathcal{T}_n)$. Furthermore, we can write the histopolant s_f in the cardinal form

$$s_f(x) = \sum_{j=1}^{n} \lambda_{\tau_j}(f)\ell_j(x). \tag{12}$$

We can transfer this basis and the histopolant s_f into the n-dimensional subspace

$$\mathcal{N}_K(\mathcal{T}, \mathcal{T}_n) = \operatorname{span}\{K(\cdot, \tau_i) \ : \ i \in \{1, \dots, n\}\}$$

of the RKHS $\mathcal{N}_K(\mathcal{T})$ by using the isometry $\Lambda : \mathcal{N}_A(\Omega) \to \mathcal{N}_K(\mathcal{T})$ introduced in Theorem 1. The respective (nodal) Lagrange basis of the space $\mathcal{N}_K(\mathcal{T}, \mathcal{T}_n)$ is given as $\{\Lambda\ell_i(\tau) \ : \ i \in \{1, \dots, n\}\}$ and the mapped histopolant in his cardinal form reads as

$$\Lambda s_f(\tau) = \sum_{j=1}^{n} \lambda_{\tau_j}(f)\Lambda\ell_j(\tau).$$

Note that if the given function f lies in the AKHS $\mathcal{N}_A(\Omega)$, we could as well write $\Lambda f(\tau_j)$ instead of $\lambda_{\tau_j}(f)$. For the reproducing kernel K, we can define the power function $P_{K,\mathcal{T}_n}(\tau)$ as

$$P_{K,\mathcal{T}_n}(\tau) := \left\| K(\cdot, \tau) - \sum_{j=1}^{n} \lambda_\tau(\ell_j)K(\cdot, \tau_j) \right\|_{\mathcal{N}_K(\mathcal{T})} = \left\| A(\cdot, \tau) - \sum_{j=1}^{n} \lambda_\tau(\ell_j)A(\cdot, \tau_j) \right\|_{\mathcal{H}},$$

where the second equality follows again by the isometry between the spaces $\mathcal{N}_A(\Omega)$ and $\mathcal{N}_K(\mathcal{T})$.

For RKHS's, the power function is a useful tool to obtain generic error estimates. The isometry in Theorem 1 allows to transfer these estimates also for AKHS's.

Corollary 1 *Assume that the averaging functionals $\{\lambda_\tau \ : \ \tau \in \mathcal{T}\}$ are continuous in $\mathcal{H}$ and generate an AKHS $\mathcal{N}_A(\Omega)$. Further, let $f \in \mathcal{N}_A(\Omega)$, $\mathcal{T}_n = \{\tau_1, \dots, \tau_n\} \subseteq \mathcal{T}$, and s_f*

the histopolant calculated upon the linearly independent average functionals $\lambda_{\tau_i}(f)$ based on the domains ω_{τ_i}, $i \in \{1, \dots, n\}$. Then, we have

$$\sup_{\tau \in \mathcal{T}} |\lambda_\tau(f - s_f)| \leq \sup_{\tau \in \mathcal{T}} P_{K,\mathcal{T}_n}(\tau) \, \|f\|_{\mathcal{H}}.$$

Corollary 1 enables us to leverage the extensive literature on (nodal) kernel interpolation and to exploit known estimates of the power function. In this context, uniform error estimates of the error $f - s_f$ are typically achieved by bounding the power function in terms of the fill distance of the set of interpolation nodes. For uniform radial averaging kernels $A(x, y) = \alpha(\|x - y\|_2)$, similar estimates for the power function $P_{K,\mathcal{T}_n}(y)$ of the reproducing kernel K can be achieved using the fill distance

$$h = h_{\Omega,\mathcal{T}_n} := \sup_{y \in \Omega} \min_{i \in \{1,\dots,n\}} \|y - y_i\|_2.$$

of the centers y_i of the domains $\omega_{y_i} = \omega + y_i$. For radial kernels K typical estimates of the power function lead to bounds of the form

$$P_{K,\mathcal{T}_n}(y)^2 \leq F(h),$$

with an increasing function $F : [0, \infty) \to [0, \infty)$ satisfying $F(0) = 0$ (Schaback, 1995).

Example 3 We consider as a radial univariate kernel the Mexican hat wavelet introduced in Example 1 (ii). The corresponding reproducing kernel K is determined in terms of a linear combination $\kappa(x) = \frac{\lambda}{2a^2}(2e^{-\frac{x^2}{\lambda}} - e^{-\frac{(x+a)^2}{\lambda}} - e^{-\frac{(x-a)^2}{\lambda}})$ of three Gaussians. By (Madych & Nelson, 1992), we know that the respective power function can be bounded by $P_{K,\mathcal{T}_n}(y) \leq e^{-\frac{\delta}{h}}$ for some $\delta > 0$. Therefore, Corollary 1 leads to the error estimate

$$\sup_{y \in \Omega} \left| \frac{1}{a} \int_{y-a/2}^{y+a/2} (f(x) - s_f(x)) \mathrm{d}x \right| \leq \|f\|_{\mathcal{N}_\Phi(\mathbb{R})} e^{-\frac{\delta}{h}}$$

for histopolation with the averaging kernel constructed upon the Mexican hat wavelet ϕ and a set $\mathcal{T}_n$ of n uniform domains with centers y_i in Ω and fill distance h between the centers.

6 Numerical Experiments: A 2D application to biomedical imaging

We now present some numerical tests that show the applicability of kernel histopolation. In these tests, we do not optimize the shapes of the kernels, although we note that procedures such as the one described in (Fasshauer & Zhang, 2007) can be applied with only minor modifications.

A natural application of two-dimensional histopolation arises in imaging. Each pixel of a grayscale image can be interpreted as the average value of a bivariate function over a small square region. From this perspective, possible applications of histopolation are image upscaling or compression.

As an example, one may compress an image by combining neighboring pixels to a larger pixel and using the average value as a new pixel information (this is also known as pixel binning). Afterwards, histopolation can be applied to the resulting low-resolution object, re-obtaining a new image with the old dimensions. In Fig. 5, we applied this procedure to a 256×256 image and compared the original to the reconstructions obtained by histopolation. Due to the rectangular alignment of the pixels, we used the tensor-product of two univariate averaging Matérn kernels with parameter $\lambda = 1$ as underlying bivariate histopolation kernel.

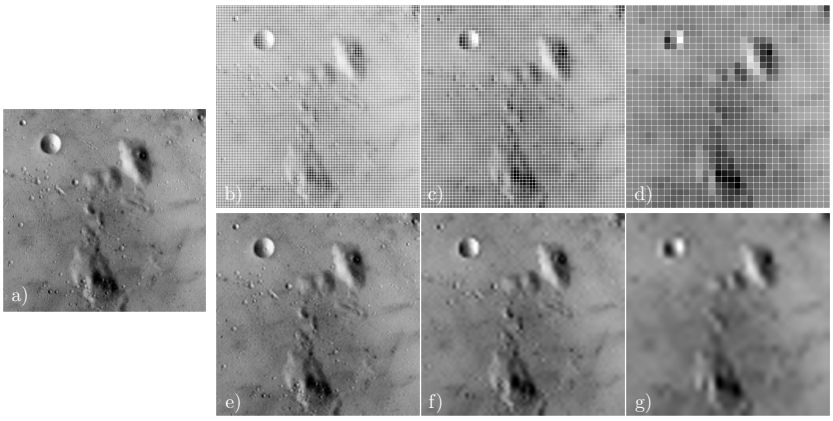


Figure 5: a) Original image of size 256×256; b)c)d) reduced image information based on 128×128, 64×64 and 32×32 average pixel values; e)f)g) respective reconstructions based on kernel histopolation using the product Matérn kernel with parameter $\lambda = 1$.

Particularly in biomedical imaging, histopolation techniques prove to be effective. In Fig. 6, we apply the histopolation algorithm to upscale an fMRI image of size 140×140, using the pixel values as averaging information. The resulting upscaled images are of size 280×280 and 420×420. As in the previous example, the histopolant is calculated using an averaged product Matérn kernel with $\lambda = 1$.

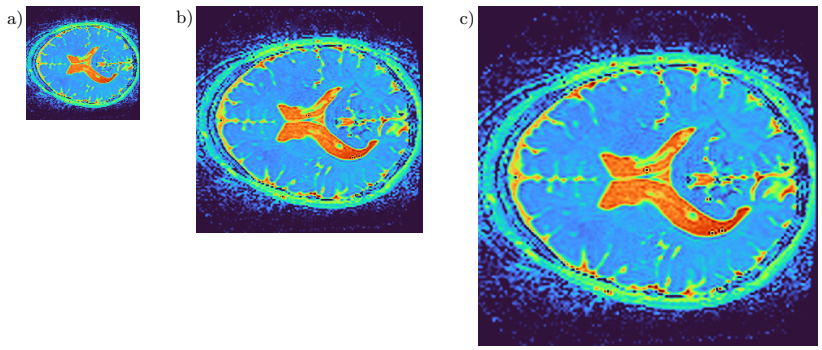


Figure 6: a) Original fMRI image of size 140×140, b) c) upscaled images of size 280×280 and 420×420 pixels using histopolation with the averaged product Matérn kernel and parameter $\lambda = 1$.

References

Aboiyar, T., Georgoulis, E. H. & Iske, A. (2010), *Adaptive ADER methods using kernel-based polyharmonic spline WENO reconstruction.* SIAM Journal on Scientific Computing, 32, 3251–3277.

Albrecht, K. (2024), *Kernel-based generalized interpolation and its application to computerized tomography.* Ph.D. thesis, University of Hamburg.

Albrecht, K. & Iske, A. (2025), *On the convergence of generalized kernel-based interpolation by greedy data selection algorithms..* Bit Numerical Mathematics, 65.

Bruni Bruno, L., Dell'Accio, F., Erb, W. & Nudo, F. (2025), *Polynomial histopolation on mock-Chebyshev segments.* Journal of Scientific Computing, 104.

Bruni Bruno, L. & Erb, W. (2024), *Polynomial interpolation of function averages on interval segments.* SIAM J. Numer. Anal., 62, 1759–1781.

Bruni Bruno, L. & Erb, W. (2025), *The Fekete problem in segmental polynomial interpolation.* BIT Numerical Mathematics, 65.

de Boor, C. (2001), "A Practical Guide to Splines". Applied Mathematical Sciences, vol. 27, revised edn. New York: Springer.

De Marchi, S., Iske, A. & Sironi, A. (2016), *Kernel-based image reconstruction from scattered Radon data.* Dolomites Research Notes on Approximation, 9, 19–31.

De Marchi, S., Iske, A. & Santin, G. (2018), *Image reconstruction from scattered radon data by weighted positive definite kernel functions.* Calcolo, 55, 399–418.

Delvos, F.-J. (1987), *Periodic area matching interpolation.* Numerical Methods of Approximation Theory: Workshop on Numerical Methods of Approximation Theory Oberwolfach, September 28 - October 4, 1986 (L. Collatz, G. Meinardus & G. Nürnberger eds). International Series of Numerical Mathematics, vol. 81. Basel: Birkhäuser Basel, pp. 54–66.

Fasshauer, G. E. & Zhang, J. G. (2007), *On choosing "optimal" shape parameters for RBF approximation.* Numer. Algorithms, 45, 345–368.

Fischer, M. & Oja, P. (2007), *Convergence rate of rational spline histopolation.* Mathematical Modelling and Analysis, 12, 29–38.

Iske, A. (1995), *Reconstruction of functions from generalized Hermite-Birkhoff data.* Approximation Theory VIII, Vol. 1: Approximation and Interpolation (C. K. Chui & L. L. Schumaker eds). World Scientific, Singapure, pp. 257–264.

Iske, A. & Sonar, T. (1996), *On the structure of function spaces in optimal recovery of point functionals for ENO-schemes by radial basis functions.* Numerische Mathematik, 74, 177–201.

Madych, W. R. & Nelson, S., *Titoloarticolo.* GiornaleAnnoPagine.

Robidoux, N. (2006), *Polynomial histopolation, superconvergent degrees of freedom, and pseudospectral discrete hodge operators.* Technical Report, 2006.

Schaback, R. (1995), *Error estimates and condition numbers for radial basis function interpolation.* Advances in Computational Mathematics, 3, 251–264.

Schoenberg, I. J. (1973), *Splines and histograms.* Spline Functions and Approximation Theory (A. Meir & A. Sharma eds). International Series of Numerical Mathematics, vol. 21. Basel: Birkhäuser Verlag, pp. 277–327.

Siewer, R. (2008), *Histopolating splines.* Journal of Computational and Applied Mathematics, 220, 661–673.

Sironi, A. (2011), *Medical Image Reconstruction Using Kernel Based Methods.* Master's thesis, University of Padova, arxiv:1111.5844v1.

Sonar, T. (1996), *Optimal recovery using thin plate splines in finite volume methods for the numerical solution of hyperbolic conservation laws.* IMA J. Numer. Anal., 16, 549–581.

Sonar, T. (1997), *On the construction of essentially non-oscillatory finite volume approximations to hyperbolic conservation laws on general triangulations: polynomial recovery, accuracy and stencil selection.* Comput. Methods Appl. Mech. Engrg., 140, 157–181.

Sonar, T. (1998), *On families of pointwise optimal finite volume ENO approximations.* SIAM J. Numer. Anal., 35, 2350–2369.

Subbotin, Y. N. (1977), *Extremal problems of functional interpolation, and mean interpolation splines.* Proceedings of the Steklov Institute of Mathematics, 138, 127—185.

van den Boogaart, K. G., Hielscher, R., Prestin, J. & Schaeben, H. (2007), *Kernel-based methods for inversion of the Radon transform on $SO(3)$ and their applications to texture analysis.* Journal of Computational and Applied Mathematics, 199, 122–140.

Wendland, H. (2004), "Scattered Data Approximation". Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press.

Wendland, H. (2005), *On the convergence of a general class of finite volume methods.* SIAM J. Numer. Anal., 43, 987—1002.

Wendland, H. (2009), *Divergence-free kernel methods for approximating the Stokes problem.* SIAM J. Numer. Anal., 47, 3158–3179.

Wu, Z. (1992), *Hermite-Birkhoff interpolation of scattered data by radial basis functions.* Approximation Theory and Applications, 8, 1–10.

Counting curves using partial differential equations

David Klompenhouwer (*)

Abstract. Enumerative geometry is the art of counting geometric objects that satisfy certain geometric conditions. For example: how many plane curves, given by polynomials of degree d, pass through a fixed set of $3d-1$ distinct points in the plane? These kinds of questions have been asked since ancient times. Partial differential equations (PDEs), on the other hand, are a much more recent tool and are used to model all sorts of physical phenomena. In the early 1990s, physicist Edward Witten made a striking conjecture that revealed an intimate connection between these two research areas. Since then, the love between enumerative geometry and PDEs has blossomed, giving rise to fascinating new insights in both areas.
In this seminar, I will give a friendly introduction to both of these topics, by focusing on objects and ideas that are of particular interest to me and showing how they are related. At the end, I will mention how my research attempts to explain this mysterious romance between enumerative geometry and PDEs.

1 A motivating geometric problem

The first postulate of Euclidean geometry is that a straight line can be drawn from any given point to any other given point. Or, equivalently: given 2 distinct points on the Euclidean plane, there is always a unique line passing through both of them.

A line in the Euclidean plane is a curve described by a polynomial equation $ax+by+c=0$ of degree 1. A conic is a curve described by a polynomial equation $ax^2+bxy+cy^2+dx+ey+f=0$ of degree 2. In the mid-1700s, mathematicians figured out that there is always a unique conic that passes through 5 fixed points in general position ("general position" here means that no three points lie on a line).

Answering the question "How many curves of degree d pass through an appropriate number of points in general position?" becomes surprisingly difficult for $d \geq 3$. More recently, mathematicians became interested in answering this question for curves in the

(*)Ph.D. course, Università di Padova, Dip. Matematica, via Trieste 63, I-35121 Padova, Italy. E-mail: klompenh@math.unipd.it . Seminar held on 18 June 2026.

complex projective plane $\mathbb{P}^2$, rather than the Euclidean plane. The complex projective plane can be defined as the quotient space

$$\mathbb{P}^2 := \mathbb{C}^3/\sim,$$

where the relation $\sim$ is given by

$$(x_0, x_1, x_2) \sim (\lambda x_0, \lambda x_1, \lambda x_2) \quad \text{for all } \lambda \in \mathbb{C} \setminus \{0\}.$$

Intuitively, it can be thought of as a "completion" of the complex plane $\mathbb{C}^2 \subset \mathbb{P}^2$ that ensures that parallel planes in $\mathbb{C}^2$ intersect at a line at ∞ in $\mathbb{P}^2$.

It turns out that the right question to ask in this context is

> Given $3d-1$ points in general position in $\mathbb{P}^2$, what is the number N_d of degree d rational curves passing through these points?

We know $N_1 = 1$, but a general answer was found only in 1995 by Maxim Kontsevich [Kon95]; [KM94] with the following recursive formula:

$$N_d = \sum_{\substack{d_1,d_2\geq 1\\ d_1+d_2=d}} d_1 d_2\, N_{d_1} N_{d_2} \left(d_1 d_2 \binom{3d-4}{3d_1-2} - d_1^2 \binom{3d-4}{3d_1-1} \right), \quad d \geq 2.$$

The first few values are

$$N_1 = 1,\ N_2 = 1,\ N_3 = 12,\ N_4 = 620,\ N_5 = 87304.$$

The solution to this simple question required some strikingly sophisticated machinery from modern enumerative algebraic geometry.

2 A modern enumerative recipe

A typical question in enumerative geometry is: how many geometric objects satisfy a given set of geometric conditions? The following is a modern recipe for formulating and solving such problems [Ric22], accompanied by the example from the previous section:

(a) Define a *moduli space* $\mathcal{M}$ that contains the geometric objects that we are interested in. Ideally, this should have an algebraic or analytic structure that reflects some key properties of the objects.

In our example, the geometric objects that we are interested in are rational curves of degree d inside $\mathbb{P}^2$. We can denote the set of all such curves by $M_0(\mathbb{P}^2, d)$. As for its algebraic structure, this fails to have the structure of an algebraic variety or complex manifold, and it even fails to be scheme. It is a more general space called Deligne-Mumford stack.

(b) Each geometric condition that we impose on the objects corresponds to a subspace of $\mathcal{M}$.

In our case, if we fix a point p_1 in $\mathbb{P}^2$, the rational curves in $\mathbb{P}^2$ passing through p_1 is a subspace $M_{p_1} \subset M_0(\mathbb{P}^2, d)$ of lower dimension.

(c) Imposing multiple geometric conditions corresponds to the intersection of the corresponding subspaces of $\mathcal{M}$. If we impose the "right" amount of conditions, this intersection will be a subspace of dimension zero, i.e. it will be a discrete set of points. Counting the number of these points gives the answer to our enumerative problem.

If we fix further points $p_2, \dots, p_{3d-1} \in \mathbb{P}^2$ we get corresponding subspaces $M_{p_2}, \dots, M_{p_{3d-1}}$. The intersection $M_{p_1} \cap M_{p_2} \cap \cdots \cap M_{p_{3d-1}} \subset M_0(\mathbb{P}^2, d)$ will only contain the degree d rational curves that pass through these points.

Here are some technicalities and difficulties that have been swept under the rug in the above recipe:

- If the moduli space $\mathcal{M}$ is not compact, one has to compactify it by adding degenerations of the geometric objects. The topology and structure of the resulting compact space $\overline{\mathcal{M}}$ can be hard to understand, but it is crucial to make the remaining steps work.
- Understanding the intersection of subspaces of $\mathcal{M}$ in a coherent manner requires working with the Chow ring $A^*(\mathcal{M})$ of $\mathcal{M}$.
- Most often, the final answer to the enumerative geometric problem will come as a *rational number*, rather than an integer. This is due to the fact that, in this formalism, geometric objects need to be counted with a weight corresponding to the size of their automorphism groups.

3 The moduli space of curves

In my research, the objects that I am most interested in are smooth (complex, connected, projective) curves of a given genus $g \geq 0$. The following moduli space is defined for every $g, n \geq 0$ such that $2g - 2 + n > 0$.

$$\mathcal{M}_{g,n} := \{(C, x_1, \dots, x_n) : C \text{ a curve of genus } g \text{ and } x_i \in C \text{ distinct}\}/\sim \ .$$

We have also added the data of n distinct marked points on the curve. Here $\sim$ means that we consider two marked curves to be the same if there is an isomorphism between them that preserves the marked points.

The space $\mathcal{M}_{g,n}$ is not compact, since families of smooth pointed curves can degenerate to a non-smooth curve. Happily, the stable reduction theorem [ACG11, Theorem 4.11] ensures that such families can always be made to degenerate to pointed *stable curves*. A stable curve is a (complex, connected, projective) curve with finitely many automorphisms, and with singularities that are at worst nodal. A point $p \in C$ is a nodal singularity if,

locally analytically near p, C is the union of two smooth branches crossing transversely at p. Thus the moduli space

$$\overline{\mathcal{M}}_{g,n} := \{(C, x_1, \dots, x_n) : C \text{ a stable curve of genus } g \text{ and } x_i \in C \text{ is smooth}\}/\sim$$

is compact and contains $\mathcal{M}_{g,n}$ as a dense open subset, i.e. it is a compactification of $\mathcal{M}_{g,n}$.

The study of the cohomology rings of $\mathcal{M}_{g,n}$ and $\overline{\mathcal{M}}_{g,n}$ is a rich and fascinating subject. David Mumford [Mum83] started this study by defining *tautological classes* on the moduli space. Their construction mirrors the construction of classes on Grassmannians arising as Chern classes of the tautological bundle.

An example of tautological classes on $\overline{\mathcal{M}}_{g,n}$ are the so-called ψ-classes, which we introduce now. We give the technical definition first, and an intuitive explanation later. To begin with, there is a morphism $\pi : \overline{\mathcal{M}}_{g,n+1} \to \overline{\mathcal{M}}_{g,n}$ that forgets the last marked point on every curve. It comes equipped with n sections $\sigma_i : \overline{\mathcal{M}}_{g,n} \to \overline{\mathcal{M}}_{g,n+1}$. The relative dualizing sheaf ω_π over $\overline{\mathcal{M}}_{g,n+1}$ is an analogue of the relative cotangent sheaf in the setting where there are nodal singularities. Pulling back ω_π along the sections σ_i gives rise to n line bundles $\mathcal{L}_1, \dots, \mathcal{L}_n$ over $\overline{\mathcal{M}}_{g,n}$.

$$\begin{array}{ccc} \omega_\pi & & \mathcal{L}_i := \sigma_i^* \omega_\pi \\ \downarrow & & \downarrow \\ \overline{\mathcal{M}}_{g,n+1} & \xrightarrow{\pi} & \overline{\mathcal{M}}_{g,n} \\ & \underset{\sigma_i}{\curvearrowleft} & \end{array} \quad , \quad i = 1, \dots, n,$$

The ψ-classes are defined as the Chern classes of these line bundles:

$$\psi_i := c_1(\mathcal{L}_i) \in H^2(\overline{\mathcal{M}}_{g,n}, \mathbb{Q}), \quad i = 1, \dots, n.$$

Intuitively, the line bundles $\mathcal{L}_i$ are geometric spaces that surject onto $\overline{\mathcal{M}}_{g,n}$, and the preimage of every element $(C, x_1, \dots, x_n) \in \overline{\mathcal{M}}_{g,n}$ under this surjection is the cotangent space $T^*_{x_i} C$. The class ψ_i is a measure of how much the line bundle $\mathcal{L}_i$ "twists", or how far it is from being isomorphic to the trivial bundle.

At the same time, the ψ-classes can also be thought of as subspaces of $\overline{\mathcal{M}}_{g,n}$, so it is natural to ask if we can understand how the different ψ-classes intersect with each other, as in the first section. The most general question one can ask is: how many points are in the following intersection of subspaces?

$$\langle \tau_{d_1} \tau_{d_2} \cdots \tau_{d_n} \rangle_g := \psi_1^{d_1} \cap \psi_2^{d_2} \cap \cdots \cap \psi_n^{d_n}. \tag{1}$$

Here $\psi_i^{d_i}$ means "ψ_i intersected with itself d_i times", which can be made to make sense if we work with the right level of rigour (see the second bullet point in section 2).

To make sense of the all the possible intersections (1), it is useful to package them into a *generating function*.

4 Generating functions

As Herbert Wilf [Wil06] puts it, "A generating function is a clothesline on which we hang up a sequence of numbers for display". To explain what this means, let us consider an example. Let $(F_n)_{n\geq 0}$ be the famous sequence of integers known as the Fibonacci numbers:

$$F_0 = 0, \quad F_1 = 1, \quad F_{n+2} = F_{n+1} + F_n, \ n \geq 0.$$

How can we package this sequence in a concise way? A possible answer is via the following power series in a variable x, which we call the generating function of the Fibonacci numbers.

$$F(x) := \sum_{n\geq 0} F_n \frac{x^n}{n!}.$$

The first and second derivatives are

$$F'(x) = \sum_{n\geq 0} F_{n+1} \frac{x^n}{n!}, \quad F''(x) = \sum_{n\geq 0} F_{n+2} \frac{x^n}{n!}.$$

The defining relation $F_{n+2} = F_{n+1} + F_n$ corresponds to a linear second-order homogeneous ODE with constant coefficients:

$$F''(x) - F'(x) - F(x) = 0.$$

The standard solution techniques for this kind of ODE, together with the initial conditions $F(0) = F_0 = 0$ and $F'(0) = F_1 = 1$, give the unique solution

$$F(x) = \frac{e^{\phi x} - e^{\psi x}}{\sqrt{5}},$$

where $\phi = \frac{1+\sqrt{5}}{2}$ and $\psi = \frac{1-\sqrt{5}}{2}$ are the golden ratio and its conjugate. An immediate consequence is Binet's well-known formula

$$F_n = \frac{\phi^n - \psi^n}{\sqrt{5}}.$$

This is an ideal scenario. We have a single generating function F in a single variable x that encodes all the elements of the sequence; the recursive relation between the elements corresponds to an ODE for $F(x)$; this ODE can be solved to give an explicit non-recursive expression for the elements of the sequence in terms of well-known mathematical constants. When one is dealing with collections of numbers that depend on each other in a highly nonlinear manner, there is a need for more sophisticated machinery.

5 Integrable hierarchies

Let us go back to our intersection numbers

$$\langle \tau_{d_1} \tau_{d_2} \cdots \tau_{d_n} \rangle_g := \psi_1^{d_1} \cap \psi_2^{d_2} \cap \cdots \cap \psi_n^{d_n}.$$

Physicist Edward Witten [Wit91] had the following idea. Introduce infinitely many variables $t_0, t_1, t_2, \dots$ (the analogs of the variable x in the previous section), and define the following generating function for these intersection numbers:

$$F(t_0, t_1, t_2, \dots) = \sum_{\substack{g \geq 0, n \geq 0 \\ 2g-2+n>0}} \sum_{d_1, \dots, d_n \geq 0} \langle \tau_{d_1} \tau_{d_2} \cdots \tau_{d_n} \rangle_g \, \frac{t_{d_1} t_{d_2} \cdots t_{d_n}}{|\mathrm{Aut}(d_1, d_2 \dots, d_n)|}.$$

From considerations coming from two-dimensional quantum gravity, he conjectured that F is a solution to the *KdV (Korteweg-de Vries) hierarchy* of partial differential equations. This is an infinite list of nonlinear partial differential equations in the infinitely many variables $t_0, t_1, t_2, \dots$ which looks like this:

$$(2n+1) \frac{\partial^3 F}{\partial t_n \partial t_0^2} = \frac{\partial^2 F}{\partial t_{n-1} \partial t_0} \frac{\partial^3 F}{\partial t_0^3} + 2 \frac{\partial^3 F}{\partial t_{n-1} \partial t_0^2} \frac{\partial^2 F}{\partial t_0^2} + \frac{1}{4} \frac{\partial^5 F}{\partial t_{n-1} \partial t_0^4}, \quad n \geq 1.$$

All the equations above are compatible, in the sense that if you start with a suitable initial condition, you can solve all of them in whichever order you want (this is what the word "integrable" refers to).

This conjecture was soon proved by Kontsevich [Kon92], and various other proofs have appeared [Mir07; KL07; OP09] using very different methods. What is remarkable is that one can deduce all of the intersection numbers (1) from this result, effectively solving the enumerative problem. The idea is similar to how the generating function

$$F(x) := \sum_{n \geq 0} F_n \frac{t^n}{n!} = \frac{e^{\phi x} - e^{\psi x}}{\sqrt{5}}$$

from the previous section encodes all the Fibonacci numbers.

6 Integrable hierarchies in enumerative geometry

Since Witten's conjecture, much effort has been made to understand how enumerative algebraic geometry and systems of partial differential equations talk to each other. Boris Dubrovin and Youjin Zhang [DZ01] initiated a vast program that assigns to any reasonable enumerative problem on $\overline{\mathcal{M}}_{g,n}$ a bihamiltonian system of PDEs, which was brought forward in [BPS12]. On the other hand, Alexandr Buryak [Bur15] gave an alternative construction via the geometry of the so-called double ramification cycle. These two approaches were shown to coincide in various situations [BLS25].

In my first project [KV25], I applied the construction of Buryak to a class of curves in $\overline{\mathcal{M}}_{g,n}$ that admit certain meromorphic differentials and theta characteristics. I showed that the corresponding enumerative problem is encoded by a solution of the BKP hierarchy (Kadomtsev-Petviashvili hierarchy of type B), which is another example of an integrable system of partial differential equations.

In my current project, I am exploring some intersection-theoretic properties of a related moduli space, called the space of *admissible covers of the projective line* [HM82; ACV03]. This is an interesting variation of the ideas developed in this seminar, since there is also a surprisingly nontrivial connection to the KdV hierarchy via an algebraic structure that arises in quantum mechanics.

References

[ACG11] E. Arbarello, M. Cornalba, and P.A. Griffiths, "Geometry of algebraic curves". Volume II. Vol. 268. Grundlehren der mathematischen Wissenschaften. With a contribution by Joseph Daniel Harris. Springer, Heidelberg, 2011, pp. xxx+963.

[ACV03] D. Abramovich, A. Corti, and A. Vistoli, *Twisted bundles and admissible covers*. Vol. 31.8. Special issue in honor of Steven L. Kleiman. 2003, pp. 3547—3618.

[BLS25] X. Blot, D. Lewánski, and S. Shadrin, *On the strong DR/DZ equivalence conjecture*. 2025. arXiv: 2405.12334 [math.AG].

[BPS12] A. Buryak, H. Posthuma, and S. Shadrin, *A polynomial bracket for the Dubrovin-Zhang hierarchies*. J. Differential Geom. 92.1 (2012), pp. 153–185.

[Bur15] A. Buryak, *Double ramification cycles and integrable hierarchies*. Comm. Math. Phys. 336.3 (2015), pp. 1085–1107.

[DZ01] B. Dubrovin and Y. Zhang, *Normal forms of hierarchies of integrable PDEs, Frobenius manifolds and Gromov-Witten invariants*. 2001. arXiv: math/0108160 [math.DG].

[HM82] J. Harris and D. Mumford, *On the Kodaira dimension of the moduli space of curves*. Invent. Math. 67.1 (1982). With an appendix by William Fulton, pp. 23–88.

[KL07] M.E. Kazarian and S.K. Lando, *An algebro-geometric proof of Witten's conjecture*. J. Amer. Math. Soc. 20.4 (2007), pp. 1079–1089.

[KM94] M. Kontsevich and Yu. Manin, *Gromov-Witten classes, quantum cohomology, and enumerative geometry*. Comm. Math. Phys. 164.3 (1994), pp. 525–562.

[Kon92] M. Kontsevich, *Intersection theory on the moduli space of curves and the matrix Airy function*. Comm. Math. Phys. 147.1 (1992), pp. 1–23.

[Kon95] M. Kontsevich, *Enumeration of rational curves via torus actions*. The moduli space of curves (Texel Island, 1994). Vol. 129. Progr. Math. Birkhäuser Boston, Boston, MA, 1995, pp. 335–368.

[KV25] D. Klompenhouwer and S. Velstra, *Spin refinement of moduli spaces of residueless meromorphic differentials and the BKP hierarchy*. 2025. arXiv: 2506.02540 [math.AG].

[Mir07] M. Mirzakhani, *Weil-Petersson volumes and intersection theory on the moduli space of curves*. J. Amer. Math. Soc. 20.1 (2007), pp. 1–23.

[Mum83] D. Mumford, *Towards an enumerative geometry of the moduli space of curves.* Arithmetic and geometry, Vol. II. Vol. 36. Progr. Math. Birkhäuser Boston, Boston, MA, 1983, pp. 271–328.

[OP09] A. Okounkov and R. Pandharipande, *Gromov-Witten theory, Hurwitz numbers, and matrix models.* Algebraic geometry — Seattle 2005. Part 1. Vol. 80, Part 1. Proc. Sympos. Pure Math. Amer. Math. Soc., Providence, RI, 2009, pp. 325–414.

[Ric22] A.T. Ricolfi, "An invitation to modern enumerative geometry". Vol. 3. SISSA Springer Series. Springer, Cham, 2022, pp. xv+302.

[Wil06] H.S. Wilf, "generatingfunctionology". Third. A K Peters, Ltd., Wellesley, MA, 2006, pp. x+245.

[Wit91] E. Witten, *Two-dimensional gravity and intersection theory on moduli space.* Surveys in differential geometry (Cambridge, MA, 1990). Lehigh Univ., Bethlehem, PA, 1991, pp. 24–310.