Politecnico di Torino

ScuDo
Scuola di Dottorato ~ Doctoral School
WHAT YOU ARE, TAKES YOU FAR

Doctoral Dissertation
Doctoral Program in Pure and Applied Mathematics (38^{th}cycle)

Topology & Dynamics

in higher-order systems and neural networks

By

Marco Nurisso

Supervisors:
Prof. Francesco Vaccarino
Prof. Giovanni Petri

Doctoral Examination Committee:
Prof. Ginestra Bianconi, Queen Mary University of London
Prof. Andrea Gabrielli, Università Roma Tre
Prof. Lamberto Rondoni, Politecnico di Torino

Politecnico di Torino
2026

Declaration

I hereby declare that, the contents and organization of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

Most of the results presented in this dissertation are adapted from the following articles:

- Nurisso, M., Arnaudon, A., Lucas, M., Peach, R. L., Expert, P., Vaccarino, F., & Petri, G. (2024). *A unified framework for simplicial Kuramoto models.* Chaos: An Interdisciplinary Journal of Nonlinear Science, 34(5).
- Nurisso, M., Morandini, M., Lucas, M., Vaccarino, F., Gili, T., & Petri, G. (2025). *Higher-order Laplacian renormalization.* Nature Physics, 1-8.
- Nurisso, M., Leroy, P., & Vaccarino, F. (2024). *Topological obstruction to the training of shallow ReLU neural networks.* Advances in Neural Information Processing Systems, 37, 35358-35387.
- Nurisso, M., Leroy, P., Petri, G. & Vaccarino, F. (2026). *Topology and geometry of the learning space of ReLU networks: connectivity and singularities.* International Conference on Learning Representations.

Marco Nurisso
2026

* This dissertation is presented in partial fulfillment of the requirements for **Ph.D. degree** in the Graduate School of Politecnico di Torino (ScuDo).

Ai miei genitori

Acknowledgements

Questa tesi è la conclusione del percorso di dottorato finanziato a valere sul PNRR, Missione 4, componente 2 “Dalla Ricerca all’Impresa” - Investimento 3.3 “Introduzione di dottorati innovativi che rispondono ai fabbisogni di innovazione delle imprese e promuovono l’assunzione dei ricercatori dalle imprese”, tramite il Decreto Ministeriale n. 352 del 9 aprile 2022 e cofinanziato dall’istituto CENTAI.

Abstract

This dissertation studies the role of topology and geometry in the analysis of complex systems, with a focus on higher-order dynamical models and neural networks. The thesis is divided into two main parts, which address distinct problems but share a common methodological foundation rooted in applied topology and discrete geometry.

The first part investigates dynamical processes on higher-order networks, modeled using simplicial complexes and hypergraphs. After introducing the necessary background in algebraic topology and topological signal processing, we study Kuramoto-like synchronization models defined on simplicial complexes, in which oscillators are associated with simplices of arbitrary dimension and interact via the complex's adjacency structure. The main contributions of this part consist in an analytical study of the equilibrium configurations of these models, their dependence on the strength of the interaction terms, and the stability properties of the resulting phase-locked states. In the process, particular attention is devoted to disentangling the meaning of the different interaction mechanisms appearing in the equations, clarifying their interpretation and their respective impact on collective synchronization. This analysis highlights how the topology of the underlying simplicial complex constrains the set of admissible equilibria and shapes their stability properties.

Building on this operator-based viewpoint, we introduce a renormalization-group framework for higher-order networks based on a novel class of operators, the cross-order Laplacians, which describe diffusion processes between simplices of different dimensions. Using spectral and entropic observables derived from these operators, we define higher-order notions of scale-invariance and propose an explicit higher-order Laplacian renormalization scheme. The method is validated on synthetic models with known hierarchical structure and applied to real-

world datasets, revealing order-specific scale-invariant signatures and non-trivial higher-order organization.

The second part focuses on the topology and geometry of the parameter space of ReLU neural networks trained by gradient flow. Exploiting the rescaling symmetries induced by homogeneous activations, we characterize the training-invariant parameter space as an algebraic variety constrained by conserved quantities. For shallow networks, we show that this invariant set can be disconnected, leading to a topological obstruction to training that depends on initialization and can reduce the network's effective expressivity. We derive exact conditions for disconnectedness and analytical estimates for the probability of encountering such obstructions under standard initialization schemes.

We then extend the analysis to general feedforward architectures modeled as directed acyclic graphs. The invariant set is described as an algebraic variety whose topology depends on architectural features such as bottlenecks and flow imbalance. We give a complete characterization of its connectedness properties, identify the singularities of this space corresponding to effective subnetworks, and show how regularization schemes, particularly nuclear norm regularization, promote convergence toward such configurations, providing a geometry-motivated form of structured pruning.

Contents

List of Figures

List of Tables

List of Symbols

A^*	Adjoint of the operator A
$A^\dagger$	Moore-Penrose pseudoinverse of the operator A
I_n	$n \times n$ identity matrix
S^d	d-dimensional sphere
Δ^k	Geometric k-simplex in $\mathbb{R}^{k+1}$
Im	Image of a linear function or Range of a matrix
Co	Convex cone generated by the columns of a matrix
$\cong$	Isomorphism or homeomorphism
ker	Kernel of a linear function or Null space of a matrix
$\mathbb{N}$	Set of natural numbers
$\mathbb{R}$	Real numbers
$\mathbb{R}_+$	Real positive numbers
$\mathbb{Z}$	Set of integers
$\mathcal{S}_n$	Symmetric group on the set $\{1, 2, \ldots, n\}$
$\operatorname{diag}(a_1, \ldots, a_n)$	Diagonal matrix with diagonal elements $a_1, \ldots, a_n$
$\|\cdot\|_2$	Euclidean norm

$\|\cdot\|_F$	Frobenius norm
$\odot$	Hadamard elementwise product of matrices
$\mathbb{1}$	Vector whose components are all equal to 1
$\oplus$	Orthogonal sum of vector spaces
$\sqcup$	Disjoint union of sets
$f \circ g$	Composition of functions

Back to Table of Contents

Chapter 0

Preliminaries

Contents

0.1 Elements of algebraic topology

In this part of the preliminaries, we give a brief overview of the theory of discrete exterior calculus, whose tools will largely be used throughout Chapters 1 and 2. The exposition loosely follows the book by Grady [2], together with [3], although with a different notation and a few differences in the mathematical constructions. The main goal will be to define and explore the properties of simplicial complexes, a particular class of combinatorial objects that extend the familiar notion of oriented graph to account for higher-order connections, called *simplices*.

Simplicial complexes take the role of geometric domains, combinatorial in nature, upon which one can define the discrete equivalent of differential forms, which we interpret as *signals*, and operators that act on them, like derivatives and Laplacians.

It is important to underline that discrete calculus is not just a *discretization* of differential calculus. It is possible to discretize (more precisely, triangulate) manifolds with simplicial complexes in such a way that the discrete objects converge to their continuous equivalents as the discretization becomes finer and finer. In this setting, discrete calculus can be seen as a discretization of calculus, useful for performing computations on continuous objects that, by their nature, cannot be fully described by a computer. Discrete calculus, however, can be applied to much more general situations and provides tools to work with other kinds of geometries, such as networks, where there is no notion of an underlying continuous space. This is the spirit in which the combinatorial theory of simplicial complexes has been successfully used in network science to address higher-order networks.

0.1.1 Simplicial complexes

Definition 0.1 (Simplicial complex)**.**

Let V be a finite set of vertices. A k**-simplex** σ is a subset $\{v_0, \dots, v_k\}$ of $k+1$ vertices in V. A **face** of σ is any subset obtained by removing one of its vertices, i.e. $\sigma_{-i} = \{v_0, \dots, v_{i-1}, \not{v}_i, v_{i+1}, \dots, v_k\}$ for some i. In this case, we say that σ is a **superface** of σ_{-i}.
An **abstract simplicial complex** X is a finite collection of simplices that is closed under the operation of taking faces: if $\sigma \in X$, then all faces of σ also belong to X.

We denote by n_k the number of k-simplices in X, X_k the set of k-simplices of X, and by K the smallest integer such that $n_{K+1} = 0$, called the **dimension** of the complex. A **subcomplex** of a simplicial complex X is simply a subset $A \subseteq X$ that is also a simplicial complex.

Geometric realization. By definition, an abstract simplicial complex encodes only combinatorial information and cannot be directly seen as a topological space in the standard sense, i.e. a set together with a collection of open sets closed by union and finite intersection [4].

In that regard, it is important to note that the set of subsets X that defines a simplicial complex is *not* a topology in this sense. In fact, the single vertices $\{v\}$ $\forall v \in V$ always belong

to the simplicial complex X and thus, if X were a topology, all of their unions would need to belong to X, resulting in the discrete topology that contains every subset of V.

Therefore, it is important to clarify here *in what sense* simplicial complexes provide a combinatorial notion of topology. The idea is that, given an abstract simplicial complex X, one can construct a geometric realization $|X|$ as a subspace of Euclidean space by gluing together geometric simplices, i.e. points, segments, triangles, tetrahedra, and their higher-dimensional analogs. With *geometric* k-simplex, we denote here the convex hull of $k+1$ affinely independent vectors

$$\mathrm{Conv}(\{u_0,\dots,u_k\}) = \left\{ \sum_{i=0}^{k} \theta_i u_i \,\middle|\, \sum_{i=0}^{k} \theta_i = 1, \theta_i \geq 0 \;\text{ for } i = 0,\dots,k \right\}.$$

Definition 0.2 (Geometric realization)**.**

Let X be a simplicial complex on a finite vertex set V. An injective map $f : V \to \mathbb{R}^d$ is a **geometric realization** of X if,

1. for every simplex $\sigma = \{v_0,\dots,v_k\} \in X$, the images $f(v_0),\dots,f(v_k)$ are affinely independent in $\mathbb{R}^d$.
2. For any $\sigma,\tau \in X$, $\mathrm{Conv}(f(\sigma)) \cap \mathrm{Conv}(f(\tau)) = \mathrm{Conv}(f(\sigma \cap \tau))$.

The geometric realization of X induced by f is the subspace of $\mathbb{R}^d$ obtained by taking the union of the convex hulls of the simplices in X mapped through f,

$$|X| := \bigcup_{\sigma \in X} \mathrm{Conv}(f(\sigma))$$

equipped with the subspace topology.

The first property of Definition 0.2 ensures that abstract k-simplices are mapped to k-dimensional simplices in $\mathbb{R}^d$, avoiding degeneracies arising from points being affinely dependent. The second property, instead, gives to $|X|$ the structure of a geometric simplicial complex, ensuring that intersections between two simplices are nonempty only when one is included in the other or when they share a face, and their intersection will coincide with that face.

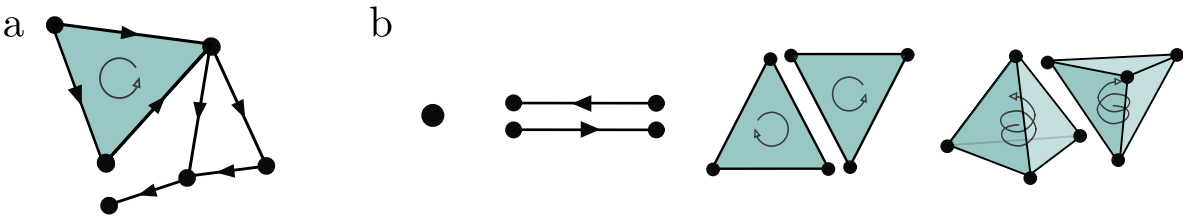

Fig. 1 **a.** A small oriented simplicial complex. **b.** Oriented simplices in their two possible orientations from order 0 to 3.

In the following, we will always work with abstract simplicial complexes, while still keeping in mind their interpretation as geometric realizations. For this reason, we will often refer to abstract 1-simplices as "edges" — , 2-simplices as "triangles" ▲, 3-simplices "tetrahedra" ◭, and so on.

Orientation. A k-simplex σ can be **oriented** by choosing an ordering of its vertices. When this is the case, we write the corresponding set of nodes using square brackets $\sigma = [v_0, \dots, v_k]$. Two orientations are considered to be equivalent when they differ by an even permutation, that is, by an even number of transpositions of elements. Hence, each simplex has exactly *two* possible orientations (see Figure 1). For instance, the oriented simplices $[a, b, c]$ and $[c, a, b]$ have the same orientation, since the latter can be obtained from the former via two swaps: $[a, b, c] \to [a, c, b] \to [c, a, b]$.

Remark 0.1 (Induced and coherent orientations)**.** *An oriented simplex naturally induces orientations on all of its faces by restricting the chosen order to the vertices of each face. For example, the oriented* 2*-simplex* $[v_0, v_1, v_2]$ *induces the orientations* $[v_0, v_1]$, $[v_0, v_2]$, *and* $[v_1, v_2]$ *on its faces. Given an oriented simplex* $\sigma = [v_0, \dots, v_k]$ *and one of its faces* $\tau = \sigma \setminus \{v_i\}$ *with the vertex* v_i *removed, we say that they are* ***coherently oriented*** $\tau \sim \sigma$ *if either* i *is even and the orientation of* τ *is equivalent to the orientation induced by* σ *or, if* i *is odd, the orientation of* τ *is* not *equivalent to the orientation induced by* σ*. Otherwise, we say that they are not coherent and write* $\tau \nsim \sigma$*.*

The role of orientations on the simplices will become clear once we start working with signals defined on top of them. For the moment, it is enough to keep in mind that orientations are not "physical" quantities, in the sense that any of the two possible orientations is a valid way to describe the *same* simplex. Orientations are best thought to provide a *reference frame* that is used to measure quantities define on top of the simplices.

Simplicial signals. Intuitively, we talk about k-dimensional signals to denote functions that associate a real number to every k-simplex, characterizing, e.g., the state of a dynamical variable at a given time. 0-dimensional signals, in particular, are functions defined on the vertices of a simplicial complex and are therefore the most natural objects that appear in graph signal processing [5]. To formalize this intuition, we start by constructing a family of linear-algebraic objects on top of the simplicial complex. We start with chain spaces, which are the most natural way to define homology groups and other aspects of the topology of simplicial complexes.

As a technical point, we note that in the following we will deal with chains and homology groups with *real* coefficients. We do this because it allows us to see the algebraic constructions through the language of real linear algebra, which may be more familiar to the reader. This choice, however, is non-standard, as most texts introduce these objects using integer coefficients. Luckily, there exist results, in particular the *universal coefficient theorem* (see [6] Chapter 3.A)), guaranteeing that one can pass from integer coefficients to real coefficients (or coefficients in any field $\mathbb{F}$) without losing any of the important aspects that we care about here.

Definition 0.3 (Chain space)**.**

For a simplicial complex X, we denote by $C_k(X;\mathbb{R})$ the real vector space generated by the oriented k-simplices of X. This space is called the **k-th chain space**, and its elements, the **k-chains**, are formal linear combinations with real coefficients of the k-simplices. By definition, changing the orientation of a simplex in a chain corresponds to multiplying it by -1.

A k-chain can therefore be written as the linear combination

$$x = \sum_{\sigma \in X_k} x_\sigma \sigma$$

with $x_\sigma \in \mathbb{R}\ \forall \sigma \in X$. The orientation property in Definition 0.3 tells us that any chain can be written as a linear combination of oriented simplices, with each simplex appearing only one time, irrespective of orientation. For instance, if $\sigma = [v_0, v_1, v_2]$, we have that

$$x_1[v_0, v_1, v_2] + x_2[v_1, v_0, v_2] = (x_1 - x_2)[v_0, v_1, v_2],$$

as $[v_0, v_1, v_2] = -[v_1, v_0, v_2]$. From this, it is clear that the set of oriented k-simplices of X forms a basis for $C_k(X;\mathbb{R})$.

From a physical point of view, the fact that the value of a signal changes sign when we change the simplex's orientation, means that simplicial chains are "*oriented quantities*". For instance, let $G = (V, E)$ be an undirected graph with nodes V and edges E. A 1-chain $x \in C_1(X; \mathbb{R})$ can be seen as an assignment of a *directed* flow to each edge. In fact, let us consider the edge $e = [v_0, v_1]$ with associated value $x_e \in \mathbb{R}$. If $x_e < 0$, we can interpret the chain as describing a flow that is judged to be negative by a reference frame that measures it going from v_0 to v_1. This, however, is equivalent to seeing the flow as positive in a reference frame that is "reversed" and measures it going from v_1 to v_0. This is captured by the relation $x_e[v_0, v_1] = -x_e[v_1, v_0]$.

It is of fundamental importance to keep this intrinsic orientation in mind when dealing with simplicial signals with the formalism outlined in this section. In fact, neglecting it can cause unexpected results that can lead to interpretative issues, as we will see in Chapter 1.

Boundary operators. A simplicial complex should be thought of as a set of objects, the simplices, which are connected by a large number of relations describing their inclusion and adjacency properties. These relations are usually formalized algebraically through the definition of a family of linear operators acting on the chain spaces.

Definition 0.4 (Boundary operator)**.**

The **boundary operator** of order k is the linear map

$$\partial_k : C_k(X; \mathbb{R}) \longrightarrow C_{k-1}(X; \mathbb{R})$$

defined on the basis elements by

$$\partial_k [v_0, \dots, v_k] = \sum_{j=0}^{k} (-1)^j [v_0, \dots, v_{j-1}, \not{v}_j, v_{j+1}, \dots, v_k],$$

and extended by linearity to k-chains.

Thus, the boundary of a k-simplex is the alternating sum of its $(k-1)$-dimensional faces, with signs chosen to take orientation into account.

A chain whose boundary is zero is called a **cycle**. Intuitively, a 1-cycle may represent a closed path of edges, while a 2-cycle can be thought of as a triangulated closed surface such as a sphere or torus. A k-chain lying in the image of ∂_{k+1} is instead called a **boundary** (Figure 2).

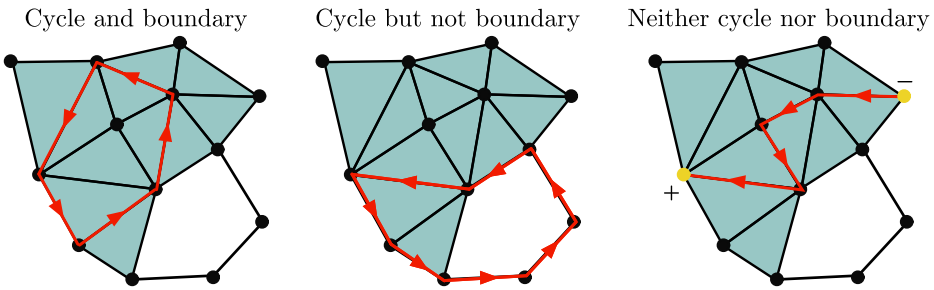

Fig. 2 (Left) In red, a 1-cycle that is also a 1-boundary. (Middle) A 1-cycle that is not a 1-boundary. (Right) A 1-chain x that is neither a cycle nor a boundary. In yellow, we represent its boundary $\partial_1 x$, which is a 0-boundary.

Chain complex. A key structural property of the chain spaces and boundary maps is that together they form a **chain complex**:

$$C_K(X;\mathbb{R}) \xrightarrow{\partial_K} C_{K-1}(X;\mathbb{R}) \xrightarrow{\partial_{K-1}} \cdots \xrightarrow{\partial_2} C_1(X;\mathbb{R}) \xrightarrow{\partial_1} C_0(X;\mathbb{R}) \xrightarrow{\partial_0} 0, \tag{1}$$

where we set $\partial_0 = 0$. The boundary operators satisfy the following fundamental property:

Proposition 0.1 (Boundary of a boundary)**.**

$$\partial_{k-1}\,\partial_k = 0 \qquad \forall\, k = 1, \dots, K. \tag{2}$$

Proof. Let us verify it for the basis k-chain representing a single k-simplex:

$$\begin{aligned}
\partial_{k-1}\partial_k[v_0,\dots,v_k] &= \sum_{j=0}^{k}(-1)^j\,\partial_{k-1}[v_0,\dots,\not{v}_j,\dots,v_k] \\
&= \sum_{j<l}(-1)^j(-1)^l[v_0,\dots,\not{v}_j,\dots,\not{v}_l,\dots,v_k] \\
&+ \sum_{j>l}(-1)^j(-1)^{l-1}[v_0,\dots,\not{v}_l,\dots,\not{v}_j,\dots,v_k] \\
&= 0,
\end{aligned}$$

as, if we swap j and l in the second summation, every term appears twice with opposite sign. □

This identity in Proposition 0.1, often summarized informally as "*the boundary of a boundary is zero,*" frequently appears in mathematics in different scenarios and has deep implications in physics, appearing in field theories of electromagnetism and gravitation [7]. At

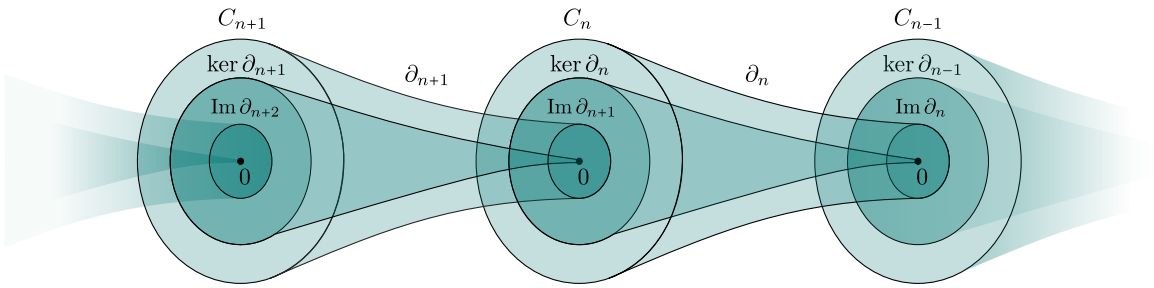

Fig. 3 Illustration of the chain complex $(C_\bullet, \partial_\bullet)$, the fundamental subspaces of the chain spaces, $\ker \partial_n$ and $\operatorname{Im} \partial_{n+1}$ and their inclusions derived from Proposition 0.1.

its core Proposition 0.1 is the linear-algebraic formulation of a purely topological fact: given a topological space X and a subspace $Y \subset X$, the boundary of the boundary of Y is always empty, $\partial\partial Y = \emptyset$.

The chain complex of Equation (1), denoted here by $(C_\bullet, \partial_\bullet)$, is a specific instance of a more general object composed of a sequence of abelian groups C_k and homomorphisms ∂_k between them with the property that $\operatorname{Im} \partial_k \subseteq \ker \partial_{k-1}$, as visualized in Figure 3.

Incidence matrices. When working in practice with these elements, we make use of the isomorphism $C_k(X, R) \cong \mathbb{R}^{n_k}$ which assigns each basis simplex to an element of the canonical basis of $\mathbb{R}^{n_k}$, and represent chains as n_k-dimensional real vectors. Accordingly, we represent the k-th order boundary operator as a matrix with the so-called *incidence matrix*. If we order the k- and $(k-1)$-simplices of X $(\sigma_j)_j$ and $(\tau_i)_i$, respectively, we can define the incidence matrix in the following way.

Definition 0.5 (Incidence matrix)**.**

We define the incidence matrix $B_k \in \mathbb{R}^{n_{k-1} \times n_k}$ as

$$B_k(i,j) = \begin{cases} 0 & \text{if } \tau_i \not\subset \sigma_j \\ 1 & \text{if } \tau_i \subset \sigma_j \text{ and } \tau_i \sim \sigma_j \\ -1 & \text{if } \tau_i \subset \sigma_j \text{ and } \tau_i \nsim \sigma_j \end{cases}$$

where, when $\tau_i \subset \sigma_j$, we write $\tau_i \sim \sigma_j$ if their orientations are coherent. For completeness, we also define $B_0 = 0 \in \mathbb{R}^{1 \times n_0}$.

Proposition 0.1 is translated to matrix form to

$$B_{k-1}B_k = 0 \;\; \forall k = 1, \dots, K. \tag{3}$$

Example 0.1. *Consider the abstract oriented simplicial complex describing a filled triangle with a segment attached,* $X = \{[v_0], [v_1], [v_2], [v_3], [v_0, v_1], [v_1, v_2], [v_0, v_2], [v_2, v_3], [v_0, v_1, v_2]\}$. *The incidence matrices are*

$$B_1 = \begin{array}{c} \\ [v_0] \\ [v_1] \\ [v_2] \\ [v_3] \end{array} \begin{array}{c} \begin{array}{cccc} [v_0,v_1] & [v_1,v_2] & [v_0,v_2] & [v_2,v_3] \end{array} \\ \begin{bmatrix} -1 & 0 & -1 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array}, \quad B_2 = \begin{array}{c} \\ [v_0,v_1] \\ [v_0,v_2] \\ [v_1,v_2] \\ [v_2,v_3] \end{array} \begin{array}{c} \begin{array}{c} [v_0,v_1,v_2] \end{array} \\ \begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \end{bmatrix} \end{array}.$$

0.1.2 Simplicial Homology

In the spirit of algebraic topology, we started with a topological object X and we constructed an *algebraic* structure, the chain complex, that captures the aspects of X we care about, namely the face relations between simplices. This mapping, as we see here, allows us to derive important topological invariants of the simplicial complex, using the language of linear algebra.

These invariants are named *homology groups* and, in our case, they are a family of vector spaces that intuitively contain information about the presence and number of holes and high-dimensional cavities in X. The algebraic formulation that we see gives an elegant, computationally tractable solution to the topological problem of formalizing what a hole is.

Definition 0.6 (Simplicial homology groups)**.**

Let X be a simplicial complex and let $(C_\bullet, \partial_\bullet)$ be its chain complex. Its k-th **homology group** with real coefficients is the vector space

$$H_k(X; \mathbb{R}) = \frac{\ker \partial_k}{\operatorname{Im} \partial_{k+1}}$$

where the quotient is intended as a quotient of vector spaces.

Note that the quotient in Definition 0.6 is well-defined because Proposition 0.1 is equivalent to stating that $\operatorname{Im} \partial_{k+1} \subseteq \ker \partial_k$ for every k.

Let us unpack Definition 0.6 and try to interpret in the light of the previous discussion. Elements of $H_k(X;\mathbb{R})$, named **homology classes**, are equivalence classes of *cycles* (elements of $\ker \partial_k$) up to the addition of *boundaries* (elements of $\operatorname{Im} \partial_{k+1}$). Two immediate consequences of this are the following:

1. all the cycles that are also boundaries belong to the null homology class $x \in \operatorname{Im} \partial_{k+1} \implies [x] = [0]$ and are said to be **null-homologous**. Boundaries, in fact, can be seen as cycles that bound filled portions of space and thus do not contain holes;

2. two cycles $x, y \in \ker \partial_k$ belong to the same homology class $[x] = [y]$ (they are **homologous**) if they differ by a boundary, i.e. there exists a $(k+1)$-chain $z \in C_{k+1}(X;\mathbb{R})$ such that $x - y = \partial_{k+1} z$. Intuitively, this means that any cycle can be "deformed" by moving part of it along a filled portion of space. In this sense, the homology group, is able to capture the topological properties of X. See Figure 4 for a visualization of this fact.

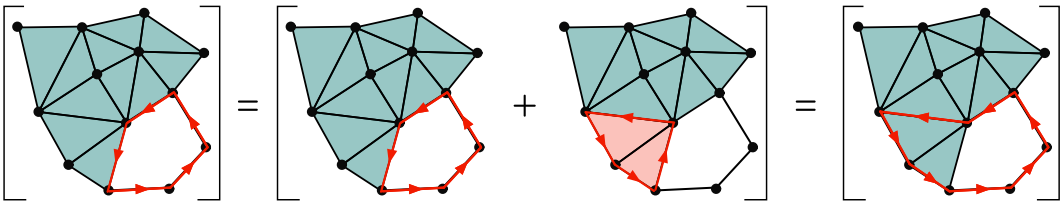

Fig. 4 A 1-cycle on a simplicial complex (left) is homologous to the same cycle with a boundary added to it (right).

Therefore, a *hole* in X is defined through the homology groups as a nonzero homology class, i.e. a cycle that is not a boundary. The number of such holes can be quantified in the following way.

Definition 0.7 (Betti numbers)**.**

The dimensions of the homology groups of X are called **Betti numbers** and are denoted by $\beta_k = \dim H_k(X;\mathbb{R})$.

Let us compute the homology groups for some simple simplicial complexes to get an idea of their meaning.

Example 0.2 (Simple homology computations)**.** • ***1.*** *For a single point $X = \{[v_0]\}$ we have the chain complex*

$$\begin{array}{ccccc} C_1 & \xrightarrow{\partial_1=0} & C_0 & \xrightarrow{\partial_0} & 0 \\ \wr\!\!| & & \wr\!\!| & & \\ 0 & & \mathbb{R} & & \end{array}$$

resulting in $H_0(X;\mathbb{R}) = \frac{\ker \partial_0}{\operatorname{Im} \partial_1} = \frac{C_0}{0} \cong \mathbb{R}$, $H_k(X;\mathbb{R}) = \begin{cases} \mathbb{R} & \text{if } k = 0 \\ 0 & \text{otherwise} \end{cases}$.

---

2. *For a single edge* $X = \{[v_0], [v_1], [v_0, v_1]\}$ *we have the chain complex*

$$\begin{array}{ccccc} C_1 & \xrightarrow{\partial_1} & C_0 & \xrightarrow{\partial_0} & 0 \\ \wr\!\!| & & \wr\!\!| & & \\ \mathbb{R} & & \mathbb{R}^2 & & \end{array}$$

with $\partial_1[v_0, v_1] = [v_1] - [v_0]$ *and* $\operatorname{rank} \partial_1 = 1, \ker \partial_1 = 0$, *resulting in* $H_0(X;\mathbb{R}) = \frac{\ker \partial_0}{\operatorname{Im} \partial_1} \cong \frac{\mathbb{R}^2}{\mathbb{R}} \cong \mathbb{R}$ *and* $H_1(X;\mathbb{R}) = \frac{\ker \partial_1}{\operatorname{Im} \partial_2} \cong 0$, $H_k(X;\mathbb{R}) = \begin{cases} \mathbb{R} & \text{if } k = 0 \\ 0 & \text{otherwise} \end{cases}$.

3. *For the disjoint union of a point and an edge* $X = \{[v_0], [v_1], [v_2], [v_1, v_2]\}$ *we have the chain complex*

$$\begin{array}{ccccc} C_1 & \xrightarrow{\partial_1} & C_0 & \xrightarrow{\partial_0} & 0 \\ \wr\!\!| & & \wr\!\!| & & \\ \mathbb{R} & & \mathbb{R}^3 & & \end{array}$$

with $\partial_1[v_1, v_2] = [v_2] - [v_1]$ *and* $\operatorname{rank} \partial_1 = 1, \dim \ker \partial_1 = 0$, *resulting in* $H_0(X;\mathbb{R}) = \frac{\ker \partial_0}{\operatorname{Im} \partial_1} \cong \frac{\mathbb{R}^3}{\mathbb{R}} \cong \mathbb{R}^2$ *and* $H_1(X;\mathbb{R}) = \frac{\ker \partial_1}{\operatorname{Im} \partial_2} \cong 0$, $H_k(X;\mathbb{R}) = \begin{cases} \mathbb{R}^2 & \text{if } k = 0 \\ 0 & \text{otherwise} \end{cases}$.

4. *For an empty triangle* $X = \{[v_0], [v_1], [v_2], [v_0, v_1], [v_1, v_2], [v_0, v_2]\}$ *we have the chain complex*

$$\begin{array}{ccccc} C_1 & \xrightarrow{\partial_1} & C_0 & \xrightarrow{\partial_0} & 0 \\ \wr\| & & \wr\| & & \\ \mathbb{R}^3 & & \mathbb{R}^3 & & \end{array}$$

Here, ∂_1 can be represented in the basis of simplices by the incidence matrix $B_1 = \begin{array}{c} \\ [v_0] \\ [v_1] \\ [v_2] \end{array} \begin{array}{c} \begin{array}{ccc} [v_0,v_1] & [v_1,v_2] & [v_0,v_2] \end{array} \\ \begin{bmatrix} -1 & 0 & -1 \\ 1 & -1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \end{array}$ *which we can use to see that* $\operatorname{rank}\partial_1 = 2$ *and* $\dim\ker\partial_1 = 1$, *resulting in* $H_0(X;\mathbb{R}) = \frac{\ker\partial_0}{\operatorname{Im}\partial_1} \cong \frac{\mathbb{R}^3}{\mathbb{R}^2} \cong \mathbb{R}$ *and* $H_1(X;\mathbb{R}) = \frac{\ker\partial_1}{\operatorname{Im}\partial_2} \cong \frac{\mathbb{R}}{0} \cong \mathbb{R}$,

$$H_k(X;\mathbb{R}) = \begin{cases} \mathbb{R} & \text{if } k = 0 \\ \mathbb{R} & \text{if } k = 1 \\ 0 & \text{otherwise} \end{cases}.$$

5. *For a filled triangle* $X = \{[v_0], [v_1], [v_2], [v_0, v_1], [v_1, v_2], [v_0, v_2], [v_0, v_1, v_2]\}$ *we have the chain complex*

$$\begin{array}{ccccccc} C_2 & \xrightarrow{\partial_2} & C_1 & \xrightarrow{\partial_1} & C_0 & \xrightarrow{\partial_0} & 0 \\ \wr\| & & \wr\| & & \wr\| & & \\ \mathbb{R} & & \mathbb{R}^3 & & \mathbb{R}^3 & & \end{array}$$

Here, ∂_1 is the same as above and ∂_2 be represented in the basis of the simplices by the incidence matrix $B_2 = \begin{array}{r} \\ [v_0,v_1] \\ [v_1,v_2] \\ [v_0,v_2] \end{array} \begin{array}{c} \begin{array}{c} [v_0,v_1,v_2] \end{array} \\ \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} \end{array}$, *which we can use to see that* $\operatorname{rank}\partial_2 = 1$ *and* $\dim\ker\partial_2 = 0$, *resulting in* $H_0(X;\mathbb{R}) = \frac{\ker\partial_0}{\operatorname{Im}\partial_1} \cong \frac{\mathbb{R}^3}{\mathbb{R}^2} \cong \mathbb{R}$, $H_1(X;\mathbb{R}) = \frac{\ker\partial_1}{\operatorname{Im}\partial_2} \cong \frac{\mathbb{R}}{\mathbb{R}} \cong 0$ *and* $H_2(X;\mathbb{R}) = \frac{\ker\partial_2}{\operatorname{Im}\partial_3} = 0$,

$$H_k(X;\mathbb{R}) = \begin{cases} \mathbb{R} & \text{if } k = 0 \\ 0 & \text{otherwise} \end{cases}.$$

The simple computations in Example 0.2 allow us to gather some important insights about the kind of information that is captured by simplicial homology groups.

First, notice that almost all the simplicial complexes examined have the 0-th homology group $H_0(X;\mathbb{R}) \cong \mathbb{R}$, except for the disjoint union of a point and an edge (Example 0.2 n.**3**) that has $H_0(X;\mathbb{R}) \cong \mathbb{R}^2$. The reason for this is that the 0-th homology group has to do with the *connectedness* of the simplicial complex X. In fact, all of the simplicial complexes described—the point, the segment, and the two triangles—are all connected, while the union of a point and an edge is disconnected and has two connected components. Hence, we can infer that the zeroth Betti number $\beta_0(X)$ computes the number of connected components of X.

Second, the only simplicial complex described that has a nonzero first homology group $H_1(X;\mathbb{R}) \neq 0$ is the empty triangle (Example 0.2 n. **4**). This is because it is the only one that contains a 1-dimensional cycle (intuitively a closed loop) that is not a boundary of a filled, two-dimensional region, i.e. a 1-dimensional hole.

These observations allow us to make the following general remark.

Remark 0.2 (Interpretation of homology groups)**.** *The k-th Betti number of a simplicial complex counts the number of k-dimensional holes present in it. In particular, β_0 counts the number of connected components ●●, β_1 the number of loops (like the one in a circle ◯), β_2 the number of 2-dimensional cavities (like the inside of a sphere ◯), and so on.*

0.1.3 Singular homology

Before going back to simplicial complexes, it is useful to introduce the construction of homology for general topological spaces, which will be used in Chapters 3 and 4 when dealing with the optimization spaces of neural networks. The singular theory mirrors the simplicial one extremely closely: one defines chains, boundary maps, a chain complex, and then homology groups.

The fundamental difference is that, as we are working with a generic topological space X, we need to find a suitable notion of simplex upon which chains can be built. This role, as we see now, is taken by *singular* simplices, defined as continuous, possibly self-intersecting (hence the name "singular") maps from the standard geometric simplex to X.

Definition 0.8 (Singular simplex)**.**

Let X be a topological space. A **singular k-simplex** in X is a continuous map

$$\sigma : \Delta^k \longrightarrow X,$$

where Δ^k is the standard geometric k-simplex in $\mathbb{R}^{k+1}$.

We denote by $C_k^{\text{sing}}(X;\mathbb{R})$ the real vector space generated by all singular k-simplices in X. Its elements are called **singular k-chains** and, just as in the simplicial case, they can be written as finite, formal linear combinations

$$x = \sum_i x_i \sigma_i, \qquad a_i \in \mathbb{R},$$

of singular k-simplices σ_i.

The boundary operator is defined as in Definition 0.4 as acting on a singular simplex by taking an alternated sum of its faces. For each $0 \leq j \leq k$, let $\delta_j : \Delta^{k-1} \hookrightarrow \Delta^k$ be the affine map that omits the j-th vertex and maps the standard $(k-1)$-simplex to the j-th face of the standard k-simplex. The **singular boundary operator**

$$\partial_k : C_k^{\text{sing}}(X;\mathbb{R}) \longrightarrow C_{k-1}^{\text{sing}}(X;\mathbb{R})$$

is defined on a singular simplex σ by

$$\partial_k \sigma = \sum_{j=0}^{k} (-1)^j \, \sigma \circ \delta_j,$$

and extended linearly to all chains. The same computation as in Proposition 0.1 shows that

$$\partial_{k-1} \, \partial_k = 0,$$

so that the sequence $(C_\bullet^{\text{sing}}(X;\mathbb{R}), \partial_\bullet)$ forms a chain complex, called the **singular chain complex** of X.

The **k-th singular homology group** is defined as the quotient

$$H_k^{\text{sing}}(X;\mathbb{R}) = \frac{\ker \partial_k}{\operatorname{Im} \partial_{k+1}}.$$

Intuitively, $H_k^{\text{sing}}(X;\mathbb{R})$ measures k-dimensional *holes* in X: closed k-dimensional surfaces (cycles) modulo those that are the boundary of a $(k+1)$-dimensional region.

The Künneth formula. Later, we will need to compute the homology groups for the Cartesian products of topological spaces. The appropriate tool to do this is given by the **Künneth formula**, which (in its simplest form) states the following.

Theorem 0.1 ([6] Corollary 3B.7)**.** *Let X, Y be topological spaces, then*

$$H_n^{\text{sing}}(X \times Y) \;\cong\; \bigoplus_{p+q=n} \Big(H_p^{\text{sing}}(X) \otimes H_q^{\text{sing}}(Y) \Big).$$

Thus, the homology of a product decomposes as a direct sum of tensor products of the homology groups of the factors. The interpretation of Theorem 0.1 becomes clear if we focus on the Betti numbers (Definition 0.7) of the product space. To do that, we employ the following important object.

Definition 0.9 (Poincaré polynomial)**.**

The **Poincaré polynomial** of a topological space X is the generating function of its Betti numbers, that is

$$p_X(t) = \sum_{k \geq 0} \beta_k t^k = \beta_0 + \beta_1 t + \beta_2 t^2 + \dots$$

Then, it is possible to prove that, as a consequence of Theorem 0.1, the Poincaré polynomial of a Cartesian product of spaces can be written as the product of the polynomials of the factors. This gives us a simple way to compute the Betti numbers of a product, given the Betti numbers of its factors.

Corollary 0.1. *Let X, Y be topological spaces and $p_X(t)$, $p_Y(t)$ be their Poincaré polynomials. Then it holds that*

$$p_{X\times Y}(t) = p_X(t) p_Y(t).$$

Example 0.3 (Torus)**.** *Let us consider, as a simple example, two circles $X = S^1$, $Y = S^1$, whose Cartesian product we know to be the torus $X \times Y = S^1 \times S^1$.*

Given that S^1 has one connected component and a single 1-dimensional hole, we have that

$$H_k(X;\mathbb{R}) = \begin{cases} \mathbb{R} & \text{if } k = 0, 1 \\ 0 & \text{otherwise} \end{cases}$$

and the Poincaré polynomial is $p_X(t) = p_Y(t) = 1 + t$. Applying Corollary 0.1, we are able to deduce that the Poincaré polynomial of the torus is

$$p_{X\times X}(t) = p_X(t)^2 = (1+t)^2 = 1 + 2t + t^2. \tag{4}$$

In Equation (4) we see that a torus has one *connected component,* two *1-dimensional holes and* one *2-dimensional hole, corresponding to its inside cavity.*

0.1.4 Homotopy equivalence

Homotopy theory provides a notion of equivalence between topological spaces that is weaker than homeomorphism and is designed to capture their global, *topological* features without looking at their local information. Intuitively, two spaces are considered the same from the viewpoint of homotopy theory if one can be continuously deformed into the other without tearing or gluing. Contracting a subspace to a lower-dimensional one, an operation that would be forbidden for a homeomorphism, is allowed in homotopy. This idea is formalized by homotopies between continuous maps and leads to the notion of homotopy equivalence, which plays a central role in algebraic topology.

Definition 0.10 (Homotopy)**.**

Let X, Y be topological spaces. Two continuous maps $f_0, f_1 : X \to Y$ are **homotopic** if there exists a continuous map $F : [0,1] \times X \to Y$ such that $F(0,\cdot) = f_0$ and $F(1,\cdot) = f_1$. F is said to be a **homotopy** from f_0 to f_1.

Definition 0.11 (Deformation retraction)**.**

Let $A \subseteq X$ be topological spaces. We say that A is a **deformation retract** of X if there exists a homotopy from the identity map of X to a **retraction** of X onto A, i.e. a continuous map $r : X \to X$ such that $r(X) = A$ and $r|_A = \mathbb{1}$.

A retraction is a special case of the more general notion of *homotopy equivalence* between topological spaces.

Definition 0.12 (Homotopy equivalence)**.**

A continuous map $f : X \to Y$ is called a **homotopy equivalence** if there exists a continuous map $g : Y \to X$ such that $f \circ g$ is homotopic to the identity on Y and $g \circ f$ is homotopic to the identity on X.

It follows that a homeomorphism is a homotopy equivalence, since one may take $g = f^{-1}$ and obtain that the compositions $f \circ g$ and $g \circ f$ are equal (and hence homotopic) to the respective identity maps.

We now provide, without proof, a central result in algebraic topology that establishes the invariance of homology to homotopy equivalence.

Theorem 0.2 (Homotopy invariance of homology, Corollary 2.11 in [6])**.** *Let X, Y be topological spaces. If there exists a homotopy equivalence between X and Y, then*

$$H_k^{\text{sing}}(X;\mathbb{R}) \cong H_k^{\text{sing}}(Y;\mathbb{R}) \; \forall k \in \mathbb{N}.$$

A particularly important special case is given by *contractible spaces*. A topological space X is said to be **contractible** if it is homotopy equivalent to a single point, or equivalently, if the identity map on X is homotopic to a constant map. Contractible spaces are, from the homotopy-theoretic perspective, topologically trivial: they contain no nontrivial holes or cycles.

As a consequence, all reduced homology groups of a contractible space vanish. In particular, if X is contractible, then $H_k(X) = 0$ for all $k > 0$, while $H_0(X) \cong \mathbb{R}$, reflecting the fact that X has a single connected component. This observation will be used repeatedly as a baseline for identifying nontrivial topological structure.

0.1.5 Relative homology

We now briefly introduce some simple elements of relative homology, a refinement of standard homology that allows us to compute the topological signatures of a simplicial complex/topological space while ignoring some of its data. As we will see in Chapter 4, this

naturally comes up when working with feed-forward neural networks, where input and output neurons need to be ignored.

In the following, we will talk about *simplicial complexes* and *subcomplexes*, with the understanding that they can be replaced by *topological spaces* and *subspaces* with some care[1].

Let us consider a simplicial complex X and a nonempty subcomplex $A \subseteq X$. We define the **relative chain space** of the pair (X, A) as the quotient space

$$C_k(X, A) = \frac{C_k(X; \mathbb{R})}{C_k(A; \mathbb{R})},$$

which is well-defined due to the natural inclusion $C_k(A; \mathbb{R}) \subseteq C_k(X; \mathbb{R})$ that follows from every simplex in A being also a simplex in X. The elements of $C_k(X, A)$, note, are equivalence classes of chains in X such that two chains are equivalent if their difference is a chain in $C_k(A)$. Therefore, with this construction, we are effectively "ignoring" the parts of the signal that are defined on A.

The boundary operator (Definition 0.4) ∂_k naturally induces a quotient boundary operator on the relative chain spaces, $\partial : C_k(X, A) \to C_{k-1}(X, A)$ giving us a chain complex

$$\cdots \to C_{k+1}(X, A) \xrightarrow{\partial_{k+1}} C_k(X, A) \xrightarrow{\partial_k} C_{k-1}(X, A) \to \dots$$

where, once again, $\partial_k \circ \partial_{k+1} = 0$. This allows us to define yet another homology group.

Definition 0.13 (Relative homology group)**.**

The **relative homology group** of the pair (X, A) is the quotient space

$$H_k(X, A) = \frac{\ker(\partial_k : C_k(X, A) \to C_{k-1}(X, A))}{\mathrm{Im}(\partial_{k+1} : C_{k+1}(X, A) \to C_k(X, A))}.$$

To get some intuition about what relative homology is computing, we can state an interesting result that shows how relative homology can be written in terms of standard homology of a different space.

Before, we need to briefly introduce the concept of **reduced homology**. Reduced homology is a slight modification of homology that is sometimes convenient. The idea is that one modifies

[1] In general, we require (X, A) to be a **good pair**, i.e. such that A is a deformation retract of some neighborhood in X. See Proposition A.5 in [6] for a proof that, if A is a nonempty subcomplex of a simplicial complex X, then (X, A) is a good pair.

the "end" of the chain complex by adding an extra $\mathbb{R}$ term in the following way:

$$\cdots \longrightarrow C_2(X;\mathbb{R}) \xrightarrow{\partial_2} C_1(X;\mathbb{R}) \xrightarrow{\partial_1} C_0(X;\mathbb{R}) \xrightarrow{\varepsilon} \mathbb{R} \longrightarrow 0, \tag{5}$$

where $\varepsilon(\sum_\sigma x_\sigma \sigma) = \sum_\sigma x_\sigma \in \mathbb{R}$. In this way, we have that $\varepsilon \circ \partial_1 = 0$ (and thus Equation (5) is still a chain complex) and, when computing reduced homology $\tilde{H}(X)$, the only group that changes is the one of order zero

$$\tilde{H}_k(X;\mathbb{R}) = \begin{cases} \frac{\ker \varepsilon}{\operatorname{Im} \partial_1} & \text{if } k = 0 \\ H_k(X;\mathbb{R}) & \text{otherwise} \end{cases}.$$

In practice, given that we are working with real coefficients, things are simplified and we have that

$$\tilde{H}_0(X;\mathbb{R}) \cong \frac{\mathbb{R}^{n_0-1}}{\operatorname{Im} \partial_1} \cong \frac{H_0(X;\mathbb{R})}{\mathbb{R}}$$

and thus the only difference is that the dimension of the reduced homology group of order 0 is equal to the dimension of the standard homology group of order 0, *decreased by one*.

We are now ready to formulate a characterization of simplicial relative homology that will be useful in Chapter 4.

Proposition 0.2 (Proposition 2.22 in [6])**.** *Let $A \subseteq X$ be a nonempty subcomplex of the simplicial complex X, then*

$$H_k(X, A) \cong \tilde{H}_k(X/A) \;\; \forall k \geq 0.$$

Therefore, Proposition 0.2 tells us that relative homology can be seen as the standard homology of a quotient space, where the subcomplex A is glued together as a single point. Doing this operation, note, can result in X/A losing the structure of a simplicial complex, and thus a more refined, technical version of simplicial homology (dealing with CW complexes) is required. Here, we refrain from getting into the details and just mention that most of the constructions and interpretations are virtually equivalent. We refer the interested reader to Chapter 2 of Hatcher's book [6] for a complete exposition.

Example 0.4 (Feed-forward neural network)**.** *As a clarification of the concepts described in this section and in anticipation of Chapters 3 and 4, we give an example of relative homology applied to a simple feed-forward neural network.*

For now, a feed-forward neural network is simply a directed acyclic graph $G = (V, E)$ with a set of input nodes V_I such that they have no edges coming in and a set of output nodes V_O with no edges coming out.

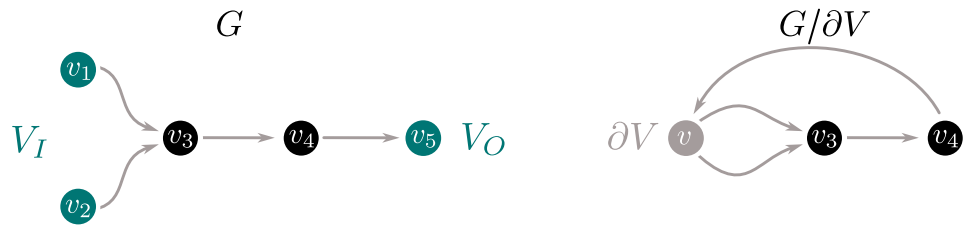

Fig. 5 **Left.** Feedforward neural network G described in Example 0.4. **Right.** The quotient graph $G/\partial V$.

Let us consider the simple network $G = (V, E)$ depicted in Figure 5 with input nodes $V_I = \{[v_1], [v_2]\}$, output nodes $V_O = \{[v_5]\}$ and edges oriented according to their direction $E = \{[v_1, v_3], [v_2, v_3], [v_3, v_4], [v_4, v_5]\}$.

Let us interpret G as a 1-dimensional oriented simplicial complex, define $\partial V = V_I \cup V_O$ the set of input and output nodes (which we interpret as a subcomplex), and let us compute the relative homology groups of the pair $(G, \partial V)$.

The relative chain complex of the pair $(G, \partial V)$ is given by:

$$\begin{array}{ccc} C_1(G, \partial V) & \xrightarrow{\partial_1} & C_0(G, \partial V) \\ \| & & \| \\ C_1(G; \mathbb{R}) & \xrightarrow{\partial_1} & C_0(G; \mathbb{R})/C_0(\partial V; \mathbb{R}) \\ | & & \| \\ \text{span}\{[1,3],[2,3],[3,4],[4,5]\} & \xrightarrow{\partial_1} & \text{span}\{[v_3],[v_4]\} \end{array}$$

where $C_1(G, \partial V) = C_1(G; \mathbb{R})$ as the subcomplex of input-output nodes ∂V contains no edges, and we are identifying the quotient $C_0(G; \mathbb{R})/C_0(\partial V; \mathbb{R})$ with the subspace of $C_0(G; \mathbb{R})$ generated by the nodes not *contained in ∂V, i.e.* span $\{[v_3], [v_4]\}$. *The boundary operator ∂_1 acts on edges mapping them to the relative chain space $C_0(G, \partial V)$ in the following way:*

$$\partial_1[v_1, v_3] = [v_3] - [v_1] = [v_3], \qquad \partial_1[v_2, v_3] = [v_3] - [v_2] = [v_3], \tag{6}$$

$$\partial_1[v_3, v_4] = [v_4] - [v_3], \qquad \partial_1[v_4, v_5] = [v_5] - [v_4] = -[v_4] \tag{7}$$

because in the quotient space we are identifying the nodes in ∂V, i.e. $[v_1], [v_2], [v_5]$, with 0*. Therefore, we can represent ∂V with the* relative *boundary matrix obtained from B_1 by removing the rows associated to the nodes in ∂V.*

$$B_1 = \begin{array}{c} \\ [v_1] \\ [v_2] \\ [v_3] \\ [v_4] \\ [v_5] \end{array} \begin{array}{c} \begin{array}{cccc} [v_1,v_3] & [v_2,v_3] & [v_3,v_4] & [v_4,v_5] \end{array} \\ \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 1 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array} \implies \tilde{B}_1 = \begin{array}{c} \\ [v_3] \\ [v_4] \end{array} \begin{array}{c} \begin{array}{cccc} [v_1,v_3] & [v_2,v_3] & [v_3,v_4] & [v_4,v_5] \end{array} \\ \begin{bmatrix} 1 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \end{array}$$

From here it is clear that

$$H_0(G, \partial V) = \frac{C_0(G, \partial V)}{\operatorname{Im} \partial_1} \cong \frac{\mathbb{R}^2}{\mathbb{R}^{\operatorname{rank} \tilde{B}_1}} \cong \frac{\mathbb{R}^2}{\mathbb{R}^2} \cong 0; \tag{8}$$

$$H_1(G, \partial V) = \frac{\ker \partial_1}{\operatorname{Im} \partial_2} \cong \ker \partial_1 \cong \mathbb{R}^{\dim \ker \tilde{B}_1} \cong \mathbb{R}^2. \tag{9}$$

Alternatively, we can apply Proposition 0.2 to see that $H_1(G, \partial V) \cong H_1(G/\partial V)$, where $G/\partial V$, represented in Figure 5, is a multigraph *obtained by merging input and output nodes and contains exactly* two *1-dimensional cycles.*

We can interpret the presence of two *relative 1-dimensional holes by noticing that there are* two *independent paths from inputs V_I to outputs V_O. These two paths, in fact, become cycles when we merge inputs with outputs in the quotient $G/\partial V$.*

To summarize, this shows how we can use relative homology to count independent paths from a given source to a given destination.

0.2 Elements of topological signal processing

0.2.1 Topological signals and the cochain complex

In discrete topology, chains represent oriented geometric domains built from simplices. Topological *signals*, however, are more naturally thought of as real-valued functions *supported on* simplices. This viewpoint is captured algebraically by a dualization of the chain complex.

The resulting objects, known as *cochains*, play the analogous role of discrete differential forms and serve as the natural setting for topological signal processing.

We note that, in general, passing from chains to cochains and from homology to cohomology leads to vastly different objects that allow topologists to derive different results. As it often happens in this thesis, we only care about finite dimensional simplicial complexes and algebraic structures with real coefficients, resulting in everything being greatly simplified.

Cochains.

Definition 0.14 (Cochain space)**.**

Let $C_k(X;\mathbb{R})$ denote the space of k-chains. The k-th **cochain space** is defined as its dual,

$$C^k(X;\mathbb{R}) := \mathrm{Hom}(C_k(X;\mathbb{R}),\mathbb{R}) = \{x : C_k(X;\mathbb{R}) \to \mathbb{R} |\ x \text{ linear}\},$$

whose elements are real functions from the chain space to $\mathbb{R}$.

Since every finite-dimensional vector space has the same dimension of its dual and is thus isomorphic to it, chains and cochains are naturally identified by taking as a basis of $C^k(X;\mathbb{R})$ the cochains that are supported on the single k-simplices. For every k-simplex $\sigma \in X_k$, we define the basis cochain

$$\hat{\sigma} : C_k(X;\mathbb{R}) \to \mathbb{R},\ \hat{\sigma}(\tau) = \begin{cases} 1 & \text{if } \tau = \sigma \\ 0 & \text{if } \tau \neq \sigma \end{cases}, \tag{10}$$

and write any cochain as

$$x = \sum_{\sigma \in X_k} x_\sigma \hat{\sigma}.$$

Coboundary operator and cohomology. Once the chain spaces are dualized, we can also dualize the boundary operator (Definition 0.4) by taking its transpose. This yields the *coboundary operator* which naturally acts in "reverse", sending a k-cochain to a $(k+1)$-cochain.

Definition 0.15 (Coboundary operator)**.**

The k-th **coboundary operator** is the linear map

$$\begin{aligned} d^k : C^k(X;\mathbb{R}) &\longrightarrow C^{k+1}(X;\mathbb{R}) \\ x &\longmapsto d^k x := x \circ \partial_{k+1} \end{aligned}$$

It is not hard to check that the coboundary operator is represented in the basis of the simplices with the matrix transpose of the boundary matrix.

Definition 0.16 (Coboundary matrix)**.**

We call **coboundary matrix** the matrix $D^k \in \mathbb{R}^{n_{k+1} \times n_k}$ defined as

$$D^k := B_{k+1}^\top.$$

If cochains are to be thought of as differential forms, the coboundary operator is the analog of the differential operator. Let us get some intuition with a simple example.

Example 0.5 (Coboundary of a 0-chain)**.** *Let us consider a* 0*-cochain* $x \in C^0(X;\mathbb{R})$*. This can be interpreted as a real function on top of the vertices of* X*. If we evaluate its coboundary* $d^0 x \in C^1(X;\mathbb{R})$ *on an edge* $[v_0, v_1]$ *we see that* $(d^0 x)_{[v_0,v_1]} = x_1 - x_0$*. The coboundary of a 0-cochain* x *thus encodes on each edge the local difference in the value of* x *on its endpoints. In this sense, we can see* d^0 *as a discrete "gradient" operation on functions.*

As with chains and boundaries, we can show that the coboundary of a coboundary is zero.

Proposition 0.3 (Coboundary of a coboundary)**.** *It holds that*

$$d^{k+1} d^k = 0,$$

Proof. Let $x \in C^k(X;\mathbb{R})$ be a k-cochain.

$$d^{k+1} d^k x = (d^k x) \circ \partial_{k+2} = (x \circ \partial_{k+1}) \circ \partial_{k+2} = x \circ (\partial_{k+1} \circ \partial_{k+2}) = 0.$$

□

Proposition 0.3 allows us to build the **cochain complex**

$$C^K(X;\mathbb{R}) \xleftarrow{d^{K-1}} C^{K-1}(X;\mathbb{R}) \xleftarrow{d^{K-2}} \cdots \xleftarrow{d^1} C^1(X;\mathbb{R}) \xleftarrow{d^0} C^0(X;\mathbb{R}) \tag{11}$$

now characterized by the property $\operatorname{Im} d^k \subseteq \ker d^{k+1}$, and define the **cohomology groups**

$$H^k(X;\mathbb{R}) := \ker(d^k)\,/\operatorname{Im}(d^{k-1})$$

defined as the space of **cocycles** modulo the space of **coboundaries**.

Analogously to De Rham's theory [8], we can take a differential-geometric point of view and think of the cohomology group as measuring the obstructions to integrating discrete potentials. In fact, if $H^k(X;\mathbb{R})$ is non-trivial, it means that there exists a cocycle x that cannot be written as a coboundary $\nexists y \in C^{k-1}(X;\mathbb{R}) : d^{k-1}y = x$, or, in other terms, a closed form that is not exact.

Over the reals, it is possible to prove that cohomology groups are isomorphic to the homology groups $H_k(X;\mathbb{R})$, so both viewpoints carry the same topological information about the presence of holes in X.

Proposition 0.4 (Isomorphism between real homology and cohomology)**.** *For every finite simplicial complex X and every k, the real homology and cohomology groups*

$$H_k(X;\mathbb{R}) \qquad \textit{and} \qquad H^k(X;\mathbb{R})$$

are isomorphic.

Proof. Let $C_k = C_k(X;\mathbb{R})$ and $C^k = C^k(X;\mathbb{R})$.

For any subspace $V \subset C_k$ we denote by $V^\circ = \left\{\varphi \in C^k : \varphi(v) = 0\ \forall\, v \in V\right\}$ its *annihilator*. Duality of the boundary-coboundary operators (Definition 0.15) implies that

$$\ker d^k = (\operatorname{Im}\partial_{k+1})^\circ, \qquad \operatorname{Im} d^{k-1} = (\ker \partial_k)^\circ.$$

Using the facts that $\dim V + \dim V^\circ = \dim C_k$ and that all chain spaces are finite-dimensional, we can write

$$\begin{aligned}\dim H^k &= \dim \ker d^k - \dim \operatorname{Im} d^{k-1} \\ &= \dim(\operatorname{Im} \partial_{k+1})^\circ - \dim(\ker \partial_k)^\circ \\ &= \dim \ker \partial_k - \dim \operatorname{Im} \partial_{k+1} = \dim H_k.\end{aligned}$$

Hence H^k and H_k are finite-dimensional real vector spaces of the same dimension and are therefore isomorphic. □

From now on, we will mainly work with cochains and cohomology.

Weights as inner products. Just like it is sometimes useful to transform a graph to a *weighted* graph by assigning *weights* to its edges, it is possible to weigh a simplicial complex by assigning weights to its simplices.

The formal and elegant way to do this is to endow each chain space with a non-trivial *inner product*. We will see how this generalizes standard constructions in graph theory like the normalized and weighted Laplacian matrices.

Let us define the diagonal matrices $W_k = \operatorname{diag}\left(w_1^k, \ldots, w_{n_k}^k\right)$, where $w_i^k > 0$ is the positive weight associated to the i-th k-simplex.

Definition 0.17 (Weighted inner product)**.**

Given two k-cochains $x, y \in C^k(X;\mathbb{R})$, represented in the basis of simplices as n_k-dimensional vectors, we define their weighted inner product

$$\langle x, y\rangle_{(k)} := x^\top W_k^{-1} y$$

which induces the norm $\|x\|_{(k)} := \sqrt{\langle x, x\rangle_{(k)}}$.

With this choice, the cochains associated to the simplices $\hat{\sigma}$ form an *orthogonal* basis of the chain space, but not an *orthonormal* one as $\hat{\sigma}_i$ will have length $\frac{1}{\sqrt{w_i^k}}$.

Having introduced the inner product, we added an additional layer of structure on the cochain space that makes $(C^k(X;\mathbb{R}), \langle\cdot,\cdot\rangle_{(k)})$ a (finite dimensional) *Hilbert space* [9]. Hilbert

spaces are richly structured and well-behaved objects, where it is possible to define angles, orthogonality and lengths of vectors.

In particular, the inner product allows us to define an analogous of the boundary operator (Definition 0.4) for cochains that takes a k-cochain and returns a $(k-1)$-cochain.

Definition 0.18 (Adjoint coboundary)**.**

We call **adjoint coboundary** the operator

$$\partial_k : C^k(X;\mathbb{R}) \to C^{k-1}(X;\mathbb{R})$$

defined as the adjoint of d^{k-1}, $\partial_k = (d^{k-1})^*$, i.e. the unique operator such that for every $x \in C^{k-1}(X;\mathbb{R})$ and $y \in C^k(X;\mathbb{R})$

$$\langle d^{k-1}x, y\rangle_{(k)} = \langle x, \partial_k y\rangle_{(k-1)}.$$

From the adjointness between ∂_k and d^{k-1}, we can easily derive the matrix representation of ∂_k that allows us to work with it in practice.

$$\langle d^{k-1}x, y\rangle_{(k)} = x^\top B_k W_k^{-1} y = x^\top W_{k-1}^{-1} W_{k-1} B_k W_k^{-1} y = \langle x, \partial_k y\rangle_{(k-1)},$$

where we see that ∂_k is represented by the matrix

$$B^k = W_{k-1} B_k W_k^{-1}. \tag{12}$$

Remark 0.3 (Uniform weights)**.** *Note that if the weights assigned to the simplices are all equal to 1, then $W_k = I_{n_k}$ $\forall k$ and the matrix representation of the adjoint coboundary (Definition 0.18) is the same as the boundary, i.e. $B^k = B_k$. For this reason, when it is clear that we are working with cochains we will call the adjoint coboundary simply "boundary".*

0.2.2 Hodge Laplacian on cochains.

Armed with inner products, boundaries, and coboundaries, we are ready to define the central object in topological signal processing: the combinatorial Hodge Laplacian [10].

Definition 0.19 (Hodge Laplacian)**.**

Let X be a finite, weighted simplicial complex; the **combinatorial Hodge Laplacian** acting on cochains is the linear map

$$\mathcal{L}^k : C^k(X;\mathbb{R}) \to C^k(X;\mathbb{R})$$

defined as

$$\mathcal{L}^k := \partial^{k+1} \circ d^k + d^{k-1} \circ \partial_k,$$

represented in the basis of the simplices by the matrix

$$L^k = B^{k+1}D^k + D^{k-1}B^k.$$

The combinatorial Hodge Laplacian is a very special operator with multiple nice properties. We start by seeing how the construction we made includes the weighted and unweighted graph Laplacian as a special case.

Remark 0.4 (Graph Laplacian)**.** *Let G be an oriented weighted graph with adjacency matrix A. We can think of it as an oriented simplicial complex with edge weights W_1 given by $1/A_{ij}$ if $A_{ij} \neq 0$. Then, if $D = A\mathbb{1}$ is the degree matrix, the weighted graph Laplacian $L = D - A$ is equal to the 0-th Hodge Laplacian L^0 and is independent of the orientation chosen for the edges. In fact,*

$$L^0 = B^1D^0 = W_0B_1W_1^{-1}B_1^\top = B_1W_1^{-1}B_1^\top$$

If one puts weights on the vertices of G equal to the inverse degrees $W_0 = D^{-1}$, we see that $L^0 = D^{-1}B_1B_1^T = D^{-1}L = I - D^{-1}A$ is the standard normalized Laplacian associated with random walks.

To interpret this operator further, we can see that its associated quadratic form measures the "smoothness" of a cochain.

Remark 0.5 (Hodge Laplacian quantifies smoothness)**.** *The quadratic form $\langle x, \mathcal{L}^k x\rangle_{(k)}$ can be interpreted as measuring the* "smoothness" *of the cochain $x \in C^k(X;\mathbb{R})$. In fact, we can*

leverage the definition of adjointness to see that

$$\begin{aligned}\langle x, \mathcal{L}^k x\rangle_{(k)} &= \langle x, \partial^{k+1} d^k x\rangle_{(k)} + \langle x, d^{k-1}\partial^k x\rangle_{(k)} \\ &= \langle d^k x, d^k x\rangle_{(k+1)} + \langle \partial^k x, \partial^k x\rangle_{(k-1)} \\ &= \left\|d^k x\right\|^2_{(k+1)} + \left\|\partial^k x\right\|^2_{(k-1)}\end{aligned}$$

which can be interpreted as measuring how much x varies along $(k+1)$-simplices (first term) plus how much it varies along $(k-1)$-simplices (second term). Therefore, $\langle x, \mathcal{L}^k x\rangle_{(k)}$ is small when the function x varies smoothly from one k-simplex to the other.

Let us state some basic properties of $\mathcal{L}^k$ without proof.

Proposition 0.5 (Basic properties of the Hodge Laplacian)**.** *The following properties hold for the combinatorial Hodge Laplacian:*

1. *$\mathcal{L}^k$ is a self-adjoint operator w.r.t. the inner product $\langle\cdot,\cdot\rangle_{(k)}$, $(\mathcal{L}^k)^* = \mathcal{L}^k$. If the complex has uniform weights, this means that the Hodge Laplacian matrix is symmetric $L^k = (L^k)^\top$.*
2. *$\mathcal{L}^k$ is a positive semi-definite operator $\langle \mathcal{L}^k x, x\rangle_{(k)} \geq 0\ \forall x \in C^k(X;\mathbb{R})$.*
3. *Due to the spectral theorem for self-adjoint operators, there exists an orthonormal (w.r.t. $\langle\cdot,\cdot\rangle_{(k)}$) basis of $C^k(X;\mathbb{R})$ of eigenvectors of $\mathcal{L}^k$ and the eigenvalues of $\mathcal{L}^k$ are real and non-negative.*
4. *It holds that* $\ker \mathcal{L}^k = \ker \partial^k \cap \ker d^k$.

When working with this object in practice, properties 1,2 and 3 of Proposition 0.5 can be translated to matrix form in the following way:

1. $(L^k)^* = W_k (L^k)^\top W_k^{-1} = L^k \iff W_k L^k = (L^k)^\top W_k \iff W_k L^k$ is symmetric;
2. $x^\top (L^k)^\top W_k^{-1} x \geq 0\ \forall x \in \mathbb{R}^{n_k}$;
3. L^k can be diagonalized as $L^k = V\Lambda V^\top W_k^{-1}$, where $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_{n_k})$ is the diagonal matrix of its non-negative eigenvalues, $V = [v_1, \ldots, v_{n_k}]$ is a square matrix whose columns form a $\langle\cdot,\cdot\rangle_{(k)}$ -orthonormal basis of eigenvectors of L^k. This means that $V^\top W_k^{-1} V = I_{n_k}$.

Property 4 is of particular interest because it tells us that cochains in the kernel of the combinatorial Hodge Laplacian are the ones that are at the same time cycles and cocycles. This property will turn out to be particularly important because of its deep connection with the topology of the simplicial complex.

Definition 0.20 (Harmonic cochains)**.**

A k-cochain $x \in C^k(X;\mathbb{R})$ is said to be **harmonic** if $\mathcal{L}^k x = 0$. The space of harmonic cochains $\ker \mathcal{L}^k$ is called the k-th **harmonic space** of X.

The following **Hodge decomposition** theorem provides a fundamental structural result for signals defined on a simplicial complex. It states that any k-cochain can be uniquely decomposed into three orthogonal components: a gradient (coboundary) part arising from lower-dimensional interactions, a curl (boundary) part induced by higher-dimensional interactions, and a harmonic part capturing global, topological features of the underlying space. This decomposition generalizes the classical Helmholtz decomposition for vector fields, where a field is split into irrotational, solenoidal, and harmonic components, to arbitrary dimensions and combinatorial settings.

Theorem 0.3 (Hodge decomposition)**.** *Given a finite simplicial complex X, the k-th cochain space can be decomposed into three orthogonal subspaces:*

$$C^k(X;\mathbb{R}) = \operatorname{Im} \partial^{k+1} \oplus \ker \mathcal{L}^k \oplus \operatorname{Im} d^{k-1},$$

meaning that every k-cochain can be uniquely written as a sum of a boundary, a coboundary and a harmonic cochain, which are all orthogonal w.r.t. $\langle \cdot, \cdot \rangle_{(k)}$.

Proof. We start by decomposing the cochain space as the orthogonal sum of the harmonic space and its orthogonal and apply Property 4 in Proposition 0.5:

$$\begin{aligned} C^k(X;\mathbb{R}) &= \ker \mathcal{L}^k \oplus (\ker \mathcal{L}^k)^\perp = \ker \mathcal{L}^k \oplus (\ker \partial^k \cap \ker d^k)^\perp \\ &= \ker \mathcal{L}^k \oplus ((\ker \partial^k)^\perp + (\ker d^k)^\perp) \\ &= \ker \mathcal{L}^k \oplus (\operatorname{Im}(\partial^k)^* + \operatorname{Im}(d^k)^*) \\ &= \ker \mathcal{L}^k \oplus (\operatorname{Im} d^{k-1} + \operatorname{Im} \partial^{k+1}). \end{aligned}$$

We just need to show that $\operatorname{Im} d^{k-1} \perp \operatorname{Im} \partial^{k+1}$. Let $x \in C^{k-1}(X;\mathbb{R})$, $y \in C^{k+1}(X;\mathbb{R})$

$$\langle d^{k-1}x, \partial^{k+1}y\rangle_{(k)} = \langle d^k d^{k-1}x, y\rangle_{(k+1)} = 0$$

due to Proposition 0.3. □

We can use the Hodge decomposition to easily prove a fundamental result that establishing an isomorphism between cohomology classes and harmonic cochains. This statement, often referred to as the *discrete Hodge theorem*, asserts that every cohomology class admits a unique harmonic representative and that the space of harmonic cochains is naturally isomorphic to the cohomology of the simplicial complex with real coefficients. In direct analogy with de Rham's theorem in the smooth setting, this result provides a discrete correspondence between topological invariants of the underlying space and particular signals defined on that space.

Theorem 0.4 (Harmonic space and cohomology)**.** *Given a finite simplicial complex X, the k-th harmonic space is isomorphic to the k-th real homology and cohomology groups*

$$\ker \mathcal{L}^k \cong H^k(X;\mathbb{R}) \cong H_k(X;\mathbb{R}).$$

Proof. We can decompose the cochain space in the following way

$$C^k(X;\mathbb{R}) = \ker d^k \oplus (\ker d^k)^\perp = \ker d^k \oplus \operatorname{Im} \partial^{k+1}.$$

Matching this with the Hodge decomposition (Theorem 0.3) tells us that $\ker d^k = \ker \mathcal{L}^k \oplus \operatorname{Im} d^{k-1}$. We now quotient out $\operatorname{Im} d^{k-1}$ and get that $\ker \mathcal{L}^k$ is isomorphic to the cohomology group $H^k(X;\mathbb{R}) = \ker d^k / \operatorname{Im} d^{k-1}$. □

If we take Theorem 0.4 together with Remark 0.5, we can see that the harmonic forms that capture the topology of X are the "smoothest" possible signals on X, in the sense that they do not vary along $(k-1)$ and $(k+1)$-simplices.

Topology & Dynamics of higher-order systems

PART I

From networks to higher-order networks. Many complex systems consist of large collections of interacting units whose collective behavior cannot be understood by studying the components in isolation. Network science emerged from the understanding that local interaction rules are often not enough, and substantial information about the system can be found in the *global* structure of interactions.

By representing interacting units as nodes and their interactions as edges, networks provide a flexible and expressive language for encoding arbitrary patterns of connectivity [11]. Within this framework, a wide range of structural principles—such as small-world organization [12] and scale-freeness [13]—and dynamical phenomena—such as synchronization [14–16] and spreading processes [17]—have been systematically analyzed. A key insight of this body of work is that collective behavior is not determined solely by interaction strength or microscopic rules but can depend crucially on the architecture of the interaction network itself.

Despite these successes, standard networks suffer from a fundamental limitation. Because they are built exclusively from nodes and edges, they can encode interactions only at the *pairwise* level. In many real systems, however, interactions are intrinsically *group-based*, involving three or more units simultaneously. For example, a social tie binding a group of friends cannot be faithfully reduced to a set of independent pairwise friendships; similarly, in chemistry, a reaction may require the simultaneous presence of multiple reagents, not merely their pairwise compatibility. In such cases, decomposing group interactions into pairs obscures the logic of the underlying process and can lead to misleading conclusions.

Empirical evidence of genuinely higher-order interactions has now accumulated across disciplines, including neuroscience [18–20], social science [21, 22], and ecology [23–25]. These observations have driven a methodological shift toward *higher-order networks*, which explicitly encode interactions among arbitrary subsets of nodes [26, 27].

Higher-order interactions can be formalized using several mathematical frameworks, most notably hypergraphs and simplicial complexes [28]. Hypergraphs offer maximal expressive power by allowing arbitrary subsets of nodes to interact. This generality, however, comes at the cost of substantial computational complexity and limited analytical tractability. Simplicial complexes impose an additional structural constraint: the presence of a group interaction implies the presence of all its lower-order sub-interactions. While less flexible, this structure enables the use of tools from combinatorial topology and discrete geometry, where simplicial complexes were originally defined as simple, computationally treatable objects that can approximate more complex objects [6]. In this setting, a higher-order network can be interpreted as a graph augmented by triangles, tetrahedra, and higher-dimensional simplices, forming a geometrical

object over which notions such as derivatives, Laplacians, and diffusion can be rigorously defined [2, 29, 30].

The shift from pairwise to higher-order interactions is not merely representational. It has been shown to qualitatively alter dynamical behavior, inducing phenomena such as multistability [31] and explosive transitions [32, 33] that are absent in the analogous purely pairwise models [34]. Within this growing literature, it is useful to distinguish between two complementary perspectives. The first extends traditional dynamical systems by endowing nodes with higher-order coupling terms, studying how group interactions modulate nodal dynamics [35–39, 28, 40–44, 38, 45–49]. The second, more topological perspective (which we adopt in Chapter 1) relaxes the classical distinction between *interacting units* and *interactions*. Here, edges, triangles, and higher-dimensional simplices are promoted to dynamical entities in their own right, each carrying independent degrees of freedom [33, 50–53]. Dynamics thus can unfold across multiple interacting topological objects, rather than being limited to nodes alone.

This viewpoint provides the foundation for the first contribution of this thesis, developed in Chapter 1, where we analytically study the *higher-order topological synchronization* model proposed in Ref. [33], following the work of Ref. [54]. By allowing synchronization dynamics to occur on simplices of different dimensions, we derive the conditions under which collective topological synchronization, which is different from standard synchronization, can emerge.

Network renormalization. In statistical physics, the renormalization group (RG) provides a systematic way to relate microscopic interactions to macroscopic behavior, identifying universal properties that persist across scales [55]. Naturally, the field has seen a great effort in trying to extend RG techniques to networks [56–63]. Doing that, however, has proven to be challenging because networks typically lack a meaningful notion of distance that can be leveraged to perform a meaningful coarse-graining of its variable. In small-world systems, nodes remain only a few steps apart regardless of system size, undermining classical spatial coarse-graining [64]. Two successful approaches tackle this problem either by equipping the network with a suitable notion of geometry (e.g. hyperbolic for scale-free networks [65, 59, 61]), or, in alternative, define scales dynamically through the process of diffusion: short times probe local structure, intermediate times reveal mesoscopic organization, and long times encode global topology [63, 66–68].

In higher-order networks, the notion of scale becomes even richer. Interactions exist simultaneously at multiple orders, and different orders may organize information in fundamentally

different ways. A system may appear scale-dependent when observed through pairwise links, yet exhibit scale invariance when examined through higher-order interactions. Most existing renormalization approaches remain node-centric, treating higher-order structures as secondary constraints. A genuinely topological renormalization scheme must instead allow interactions themselves to coarse-grain and flow across scales.

This observation motivates the second contribution of this thesis, developed in Chapter 2 following Ref. [67]. There, we discuss a renormalization group framework for higher-order networks based on higher-order diffusion processes. Central to this construction is the *cross-order Laplacian*, which gives us a tool to study diffusion happening between higher-order relations through higher-order relations. This approach reveals that the effective scale of a complex system is not unique, but depends on the topological lens through which it is observed.

Back to Table of Contents

Chapter 1

A unified framework for Simplicial Kuramoto models

Contents

1.1 Introduction

Synchronization is defined as the emergence of order from the interactions among many parts. It is a ubiquitous phenomenon that occurs in both natural and human-engineered systems [14–16] and can be observed in a wide range of systems, including the firing of neurons [69], the twinkling of fireflies [70], power grids [71, 72] or audience applause [73]. Despite the complexity and differences of these systems, the canonical Kuramoto model [74] provides a unified framework for describing the onset of synchronization in systems of oscillators that interact in a pairwise fashion. While the original version of the model included interactions between all pairs of oscillators, later extensions of the model allowed the specification of arbitrary network topologies [75]. This, in turn, revealed interesting relationships between the dynamical properties of the model and the structure of the underlying network [76, 12].

Traditional networks, however, provide a limited perspective on complex systems as they only consider pairwise interactions. To overcome this limitation, a new paradigm has recently emerged: networks with group (or higher-order) interactions, i.e., interactions between any number of units [26, 77, 78]. Higher-order interactions have been studied across a variety of dynamical processes, including random walks [79], diffusion [29, 80, 81], consensus [82–84], spreading [85–91], percolation [92–94], and evolutionary game theory [95].

Naturally, the study of higher-order interactions has also reached synchronization. One way to approach the modeling of synchronization in higher-order systems is to extend the family of possible interactions to include groups. From a network of interacting oscillators, we pass to a simplicial complex where node oscillators can also interact through triangles, tetrahedra, or higher-order structures, (see for example Refs [35–39, 28, 40–44, 38, 45–49]). Another approach is to consider the *simplicial Kuramoto* [33] as a model of synchronizing dynamics of higher-order topological signals. With it, we are not constrained to consider the evolution of oscillators placed on nodes, but we can place them on simplices of any order. This change, which may initially appear arbitrary, allows us to consider higher-order interactions in a novel and powerful way: if an edge can connect only two nodes at a time, a triangle connects three edges, a tetrahedron connects four triangles, etc. More generally, simplicial oscillators of order k will interact through $(k+1)$-simplices, resulting in interactions of order $k+2$. Its

evolution equation, moreover, can be elegantly written by borrowing some of the concepts of discrete exterior calculus [2], the discrete analogous to differential geometry on manifolds. This geometric structure allows us to get precious insights into the dynamics of the model and how it is related to the topological properties of the simplicial complex. This fruitful relation with topology has also recently put the simplicial Kuramoto at the center of the attention, resulting in different variants and extensions of the original model [96–98, 50].

In this chapter, which is based on Ref. [54], we aim to lay down the mathematical foundations for the analytical study of Kuramoto models on simplicial complexes. In particular, we focus on the models belonging to the second class mentioned above, that is the ones where oscillators themselves are simplices and interact through the adjacency structure encoded in a simplicial complex.

1.2 The Kuramoto model

We begin by briefly introducing the classical Kuramoto model. Let us consider a system of n *phase* oscillators, characterized solely by their phase θ_i and **natural frequency** ω_i, the frequency at which they oscillate when isolated from any interactions. The evolution of the uncoupled system can be described by a set of differential equations: $\dot{\theta}_i = \omega_i$ for each oscillator i. To account for the interaction among oscillators, various approaches can be employed, depending on the underlying physics of the phenomenon under investigation. A particularly elegant and widely studied model, renowned for its simplicity and analytical tractability, was introduced by Kuramoto [74]. Known as the Kuramoto model, it is described by the following system of first-order differential equations:

$$\dot{\theta}_i = \omega_i - \sigma \sum_{j=1}^{n} \sin(\theta_i - \theta_j). \tag{1.1}$$

In this formulation, an additional term captures the effect of interactions between oscillator i and every other oscillator j, modulated by a positive **coupling** or *interaction strength* parameter, denoted as σ. By the properties of the sine function, we observe that the interaction term between oscillators i and j becomes zero when $\theta_i - \theta_j = k\pi$ i.e. when θ_i and θ_j represent the same or opposite angles modulo 2π. Conversely, the interaction is strongest when the phase difference between the oscillators corresponds to odd multiples of $\frac{\pi}{2}$, implying orthogonal states on the unit circle. This simple interaction mechanism forms the basis of the Kuramoto

model, whose variants and generalizations [99–101] give it the capability of having a much richer behavior.

1.3 The simplicial Kuramoto model

In this section, we formulate and study the Kuramoto model for interacting simplicial oscillators proposed in Ref. [33]. The rest of this section is organized as follows:

- In Section 1.3.1, we formulate the simplicial Kuramoto model using the tools of discrete calculus introduced in Chapter 0.
- In Section 1.3.3, we describe the local form of the two types of interactions in the model: from below and from above. In the case of interactions from below, we identify the presence of self-interactions resulting from free faces.
- In Section 1.3.4, we show that the k-th order simplicial Kuramoto model and the standard node Kuramoto model are equivalent when the simplicial complex is a simplicial k-manifold.
- In Section 1.3.5, we describe how the dynamics naturally split into three independent subdynamics using the combinatorial Hodge decomposition theorem.
- In Section 1.3.6, we recall the definition of simplicial order parameter proposed in Ref. [50], discuss its implications on the meaning of synchronization in the simplicial model, and its differences with the standard Kuramoto order parameter.

1.3.1 Simplicial Kuramoto model

Given a simplicial complex X (Definition 0.1), possibly weighted, the k-th order simplicial Kuramoto model [33, 50] describes a system where the k-simplices are oscillators interacting through common faces and higher-order simplices. For example, one can consider oscillating edges that interact through common nodes and triangles (see Figure 1.2a). The model can be elegantly formulated with the boundary and coboundary matrices (Definitions 0.5 and 0.16) in the following way.

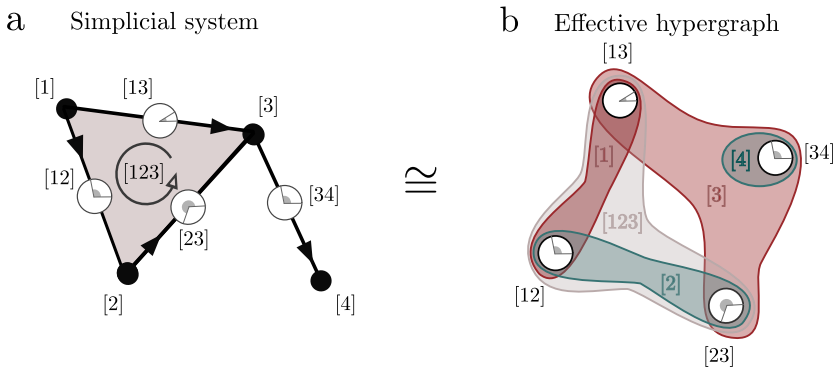

Fig. 1.1 **a.** Edge simplicial Kuramoto on the simplicial complex described in Example 1.1. **b.** The effective hypergraph of the dynamics describing the actual interactions taking place between the oscillators. The interaction hyperedges are labeled by the name of the simplex in the original complex which generates them. Note how the hyperedge $[4]$, representing the term $\sin(\theta_{[34]})$ in Equation (1.5), is interpreted here as a self-interaction.

Definition 1.1 (Simplicial Kuramoto)**.**

The k-th order simplicial Kuramoto dynamics on a simplicial complex is given by the system of ordinary differential equations

$$\dot{\theta}_{(k)} = \omega - \sigma^{\uparrow} B^{k+1} \sin\left(D^k \theta_{(k)}\right) - \sigma^{\downarrow} D^{k-1} \sin\left(B^k \theta_{(k)}\right) . \tag{1.2}$$

Here, the phases of the oscillating k-simplices are gathered in the n_k-dimensional vector $\theta_{(k)}$, formalized as a real k-cochain $\theta_{(k)} \in C^k(X;\mathbb{R})$ (Definition 0.14), while $\omega \in C^k(X;\mathbb{R})$ represents the natural frequencies, i.e. ω_i is the frequency at which oscillator i oscillates when no interactions are present. Notice that it is customary to interpret the phases to lie in the unit circle $\mathbb{R}/2\pi\mathbb{Z}$, but it is convenient to work with them as if they were real numbers. Here, the entries of ω can take any real value and are not required to be equal, meaning that we study the general case of *non-identical oscillators*, unless stated otherwise.

Finally, the parameters $\sigma^{\uparrow}, \sigma^{\downarrow} > 0$ represent, respectively, the strength of the coupling through superfaces and faces. As stated in Ref. [83], the two interaction terms $B^{k+1}\sin(D^k\theta)$ and $D^{k-1}\sin(B^k\theta)$ describe, respectively, interactions from *above* and *below*, i.e. each oscillating k-simplex interacts with its adjacent simplices through both higher $(k+1)$- and lower dimensional $(k-1)$-simplices (Figure 1.2). In Section 1.3.3, we unpack the matrix formulation and see the explicit form of these interaction terms.

For ease of notation, from now on we will drop the subscript from $\theta_{(k)}$, as the order of oscillation can be easily inferred by the indices of the boundary and coboundary matrices in Definition 1.1.

The form of Definition 1.1 is not arbitrary but comes from the fact that, for $k = 0$, it reduces to the standard Kuramoto model (from now on referred to as "node Kuramoto") on a network:

$$\dot{\theta}_i = \omega_i - \sigma \sum_j A_{ij} \sin(\theta_i - \theta_j), \tag{1.3}$$

where A is the graph adjacency matrix

Proposition 1.1. *When $k = 0$, the simplicial Kuramoto of Definition 1.1 reduces to the standard Kuramoto on the graph underlying the simplicial complex X.*

Proof. For $k = 0$, Definition 1.1 becomes

$$\dot{\theta} = \omega - \sigma B^1 \sin(D^0\theta). \tag{1.4}$$

Let $G = (V, E)$ be the unweighted graph underlying X i.e. the one containing its vertices and the 1-simplices as edges. First, compute the action of $D^0 = B_1^\top$ (Definition 0.16) on the phases vector. For any edge $\epsilon \in E$

$$(B_1^\top \theta)_\epsilon = \sum_{i=1}^{N} (B_1^\top)_{\epsilon i}\theta_i = \sum_{i=1}^{N} (B_1)_{i\epsilon}\theta_i = \theta_{\mathbf{h}(\epsilon)} - \theta_{\mathbf{t}(\epsilon)},$$

where $\mathbf{h}(\epsilon), \mathbf{t}(\epsilon)$ give respectively the head and tail node of edge ϵ. It follows that, for any node i,

$$\begin{aligned}
\left[B_1 \sin(B_1^\top \theta)\right]_i &= \sum_{\epsilon\in E} (B_1)_{i\epsilon} \sin(B_1^\top\theta)_\epsilon = \sum_{\epsilon\in E} (B_1)_{i\epsilon} \sin(\theta_{\mathbf{h}(\epsilon)} - \theta_{\mathbf{t}(\epsilon)}) \\
&= \sum_{\epsilon:\mathbf{h}(\epsilon)=i} \sin(\theta_i - \theta_{\mathbf{t}(\epsilon)}) - \sum_{\epsilon:\mathbf{t}(\epsilon)=i} \sin(\theta_{\mathbf{h}(\epsilon)} - \theta_i) \\
&= \sum_{\epsilon:\mathbf{h}(\epsilon)=i} \sin(\theta_i - \theta_{\mathbf{t}(\epsilon)}) + \sum_{\epsilon:\mathbf{t}(\epsilon)=i} \sin(\theta_i - \theta_{\mathbf{h}(\epsilon)}) \\
&= \sum_{j\in\mathcal{N}(i)} \sin(\theta_i - \theta_j) = \sum_{j=1}^{n} A_{ij} \sin(\theta_i - \theta_j),
\end{aligned}$$

where $\mathcal{N}(i)$ is the neighborhood of node i. □

One can think of D^0 as projecting the node phases on the edges by associating to each edge, which describes an interaction, the difference of its endpoints' phases. The boundary operator B^1 then projects the interactions back to the nodes, so that each node receives contributions from all edges that are incident to it.

The extension of this term to higher-order oscillators is straightforward once one sees the model as a nonlinear extension of the graph Laplacian $L^0\theta = B^1 D^0\theta \;\rightarrow B^1 \sin(D^0\theta)$ (Remark 0.4), which can be naturally generalized with the discrete Hodge Laplacian of Definition 0.19.

Notice that in the case of the node Kuramoto, no simplices with order lower than the nodes exist; hence, the dynamics results from interactions from *above*, as the left term of Definition 1.1. The interaction term from *below* is naturally introduced to account for the lower adjacency structure present in simplicial complexes, but absent in graphs. Simply put, two triangles can be adjacent through a common edge and a common tetrahedron, but two nodes can only be adjacent through an edge, i.e. a higher-order simplex. These dynamics belongs to a wider class of dynamical systems on simplicial complexes, whose stability properties have been studied when the sine is replaced with a general nonlinearity in Refs. [83, 102, 103]. Moreover, the properties of node-like phase-oscillator models with general nonlinear interaction functions have been extensively studied in Refs. [104–113].

Example 1.1. *Let us now study explicitly a simple example of the simplicial Kuramoto dynamics of the edge oscillators on a small, unweighted simplicial complex (Figure 1.1a). The simplicial complex of interest is given by* $X = \{[1], [2], [3], [4], [1,2], [1,3], [2,3], [3,4], [1,2,3]\}$ *and, taking into account the orientations, the incidence matrices of order* 1 *and* 2 *are*

$$B_1 = \begin{array}{c} \\ [1] \\ [2] \\ [3] \\ [4] \end{array} \begin{array}{c} \begin{array}{cccc} [1,2] & [1,3] & [2,3] & [3,4] \end{array} \\ \begin{bmatrix} -1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array}, \quad B_2 = \begin{array}{c} \\ [1,2] \\ [1,3] \\ [2,3] \\ [3,4] \end{array} \begin{array}{c} \begin{array}{c} [1,2,3] \end{array} \\ \begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \end{bmatrix} \end{array}.$$

Since X is unweighted, we also have that adjoint coboundary (Definition 0.18) matrices simply correspond to the incidence matrices $B^1 = B_1$, $D^0 = B_1^\top$, $B^2 = B_2$, $D^1 = B_2^\top$. If we consider the vector of phases on the edges $\theta = (\theta_{[1,2]}, \theta_{[1,3]}, \theta_{[2,3]}, \theta_{[3,4]})$ and zero natural frequencies

$\omega = 0$, then, after some algebra, we see that Definition 1.1 becomes

$$\begin{cases} \dot{\theta}_{[12]} = -\sigma^{\downarrow}\left(\sin(\theta_{[12]}+\theta_{[13]})+\sin(\theta_{[12]}-\theta_{[23]})\right) \\ \qquad\quad -\sigma^{\uparrow}\sin(\theta_{[12]}-\theta_{[13]}+\theta_{[23]}) \\ \dot{\theta}_{[13]} = -\sigma^{\downarrow}\left(\sin(\theta_{[13]}+\theta_{[12]})+\sin(\theta_{[13]}+\theta_{[23]}-\theta_{[34]})\right) \\ \qquad\quad +\sigma^{\uparrow}\sin(\theta_{[12]}-\theta_{[13]}+\theta_{[23]}) \\ \dot{\theta}_{[23]} = -\sigma^{\downarrow}\left(\sin(\theta_{[23]}-\theta_{[12]})+\sin(\theta_{[23]}+\theta_{[13]}-\theta_{[34]})\right) \\ \qquad\quad -\sigma^{\uparrow}\sin(\theta_{[12]}-\theta_{[13]}+\theta_{[23]}) \\ \dot{\theta}_{[34]} = -\sigma^{\downarrow}\left(\sin(\theta_{[34]}-\theta_{[13]}-\theta_{[23]})-\sin(\theta_{[34]})\right) \end{cases}, \tag{1.5}$$

where the interaction terms from below and from above are identifiable by their coupling strengths. Notice that some of the interactions from below are pairwise, others are higher-order (they involve three oscillators) and one, $\sin(\theta_{[3,4]})$, is of order 0 as it depends on the value of a single oscillator. Moreover, the interaction from above through the triangle, as we expected, is higher-order and involves three oscillators.

While it is natural to define the dynamics on the edges and formulate the model using the incidence matrices of the simplicial complex, it is interesting to look at Equation (1.5) from another point of view. If we forget about the underlying simplicial complex and that $\theta_{[i,j]}$ is a phase associated with an edge oscillator, what we are left with is a dynamical system where the phases of four different oscillators evolve by interacting with each other in a way specified by the functional form of the equations. It is natural, therefore, to consider these oscillators as nodes and represent their interactions with *hyperedges*, i.e. arbitrary groups of nodes. What we get, by neglecting the signs inherited by the orientations, is an *effective hypergraph* (Figure 1.1b) which does not resemble the original simplicial complex but has the advantage of clearly representing the actual interaction structure underlying the dynamics. In general, the effective hypergraph has the k-simplices as its nodes, and has one hyperedge for each $(k+1)$ and $(k-1)$-simplex responsible for the interaction. Thus, the simplicial Kuramoto model can be seen as a particular kind of hypergraph oscillator dynamics where the coupling functions depend on the orientations of the original simplices.

Notice how the coupling functions in Equation (1.5) are different from the ones classically used in hypergraph oscillator models [35, 37–39, 28, 40] as they do not possess the standard phase-shift symmetry which dictates that increasing the phase of every oscillator by the same amount should leave the dynamics unchanged. This is due to the presence in the model of a

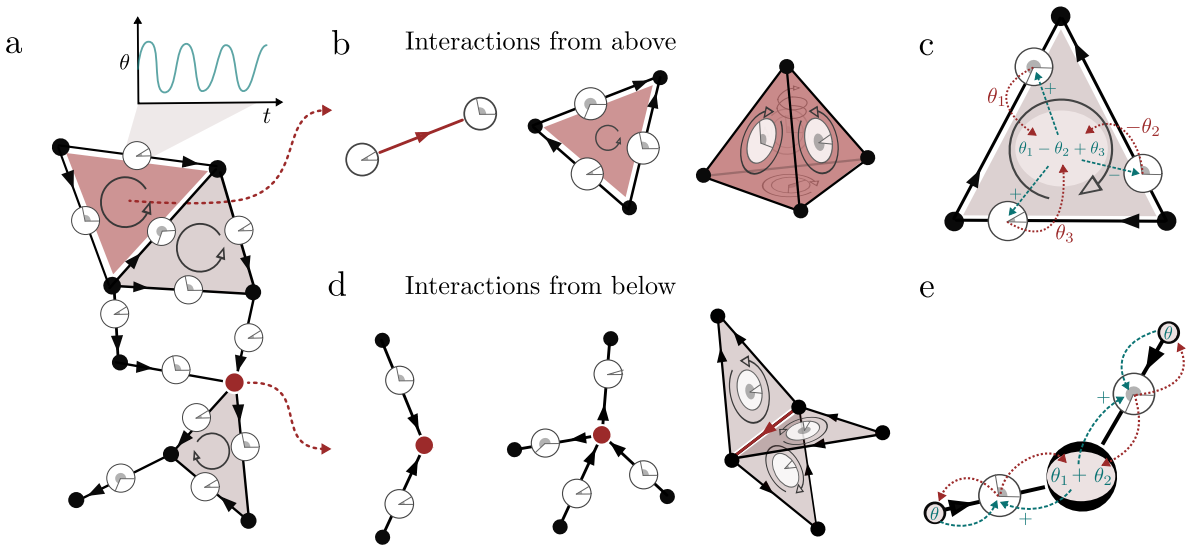

Fig. 1.2 **a.** The simplicial Kuramoto model allows us to consider oscillators, shown here as clocks, on the edges of a simplicial complex, interacting through nodes and triangles. **b.** The interaction from *above* Equation (1.11) happens between $k+2$ oscillating k-simplices through a single $(k+1)$-simplex, here highlighted in red. **c.** In the interaction from above, each oscillator involved is influenced by a term depending on the oriented sum of the phases. The phase θ_2 appears with a minus sign because its edge is oriented in the opposite direction relative to the triangle. **d.** The interaction from *below* Equation (1.14) happens between an arbitrary number of oscillating k-simplices through a single $(k-1)$-simplex, highlighted here in red. **e.** Unlike interactions from above, interactions from below through free faces are akin to self-interactions.

different symmetry, the *harmonic* phase-shift symmetry, which we discuss in more detail in Section 1.3.5.

1.3.2 Simplicial Kuramoto as a system of resonant oscillators

Interestingly, we can connect the functional form of the simplicial Kuramoto to a system of resonantly interacting oscillators, following the procedure outlined in Ref. [15, Section 8.1].

Consider first the case in which ω is close to being a k-cocycle, i.e. $D^k\omega \approx 0$. Now, we consider a single interaction from above given by the $(k+1)$-simplex $a = [a_0, \dots, a_{k+1}]$, with the oscillators on its faces indexed by $0, \dots, k+1$. Then, the condition $D^k\omega \approx 0$ can be written explicitly as

$$\xi_{a,0}\omega_0 + \xi_{a,1}\omega_1 + \cdots + \xi_{a,k+1}\omega_{k+1} \approx 0\,, \tag{1.6}$$

where $\xi_{a,j} \in \{-1, 1\}$ is the relative orientation of a w.r.t. its face with index j.

Additionally, the evolution of the oscillators' phases θ_j can be written in general form

$$\dot{\theta_j} = \omega_j + \cdots + \epsilon Q_a(\theta_0, \ldots, \theta_{k+1}) \,, \tag{1.7}$$

where Q_a is a generic interaction function between $k+2$ oscillators associated with a and the dots refer to the other interactions in which oscillator j is involved. We also make the assumption of weak coupling $\epsilon \ll 1$. The function $Q_a(\theta_0, \ldots, \theta_{k+1})$ is 2π-periodic in each of its arguments so that it can be written in Fourier series as

$$Q_a(\theta_0, \ldots, \theta_{k+1}) = \sum_{s_0,\ldots,s_{k+1} \in \mathbb{Z}} f_a^{s_0,\ldots,s_{k+1}} e^{i \sum_n s_n \theta_n} \,, \tag{1.8}$$

where $f_a^{s_0,\ldots,s_{k+1}}$ are complex coefficients indexed by the integers $s_0, \ldots, s_{k+1}$. Since the coupling is weak, we can now write the phase as $\theta_j = \omega_j t + \varphi_j$ i.e. as the contribution of a *fast phase* that results from the natural frequency and a *slow phase* φ_j, varies slowly on the time scale of ω_j^{-1}. Substituting this into Equations (1.7) and (1.8), we get the evolution of the slow phase

$$\dot{\varphi}_j = \cdots + \varepsilon \sum_{s_0,\ldots,s_{k+1} \in \mathbb{Z}} f_a^{s_0,\ldots,s_{k+1}} e^{it \sum_n s_n \omega_n} e^{i \sum_n s_n \varphi_n} \,. \tag{1.9}$$

We may now perform a time averaging over the same scale of the natural periods ω_l^{-1} of the oscillators involved. In this way, the slow phases φ_l can be assumed to be constant and, we see, all the terms ω_n in the sum with explicit time dependence will correspond to fast oscillation. These will all result in a null contribution except those that satisfy the resonance condition

$$s_0\omega_0 + s_1\omega_1 + \cdots + s_{k+1}\omega_{k+1} \approx 0 \,. \tag{1.10}$$

Given that we chose ω to be a k-cocycle (1.6), we have that Equation (1.10) is satisfied with $s_i = m\xi_{a,i}$ for any $m \in \mathbb{Z}$.

Then, assuming the only relevant resonant terms are the integer multiples of the relative orientations $(\xi_{a,0}, \ldots, \xi_{a,k+1})$, we have that

$$\begin{aligned} Q_a(\theta_0, \ldots, \theta_{k+1}) &= \sum_{m \in \mathbb{Z}} f_a^{m\xi_{a,0},\ldots,m\xi_{a,k+1}} e^{im \sum_j \xi_{a,j} \theta_j} \\ &= q\left(\sum_j \xi_{a,j} \theta_j \right) \,. \end{aligned}$$

If, moreover, $q(\psi) = \sin(\psi)$, we get the coupling of the form obtained in Ref. [114] for three resonantly interacting oscillator communities and of the simplicial Kuramoto model interactions.

This resonance condition, however, only holds for a single $k+1$ simplex a in the simplicial complex. Hence, it is then natural to ask whether it is possible to find frequencies ω which are in a similar resonance relation for all interactions at once. As the cocycle condition $D^k\omega \approx 0$ means that $(D^k\omega)_a \approx 0$ for every $(k+1)$-simplex a, the resonance condition holds for all interactions from above. For interactions from below, one can check that we instead can ask for ω to be an approximate k-cycle $B^k\omega \approx 0$. Therefore, if we have $B^k\omega \approx 0$ and $D^k\omega \approx 0$ at the same time, we will have a resonance condition which holds for all interactions in the simplicial model. Moreover, $B^k\omega \approx 0$ and $D^k\omega \approx 0 \iff L^k\omega \approx 0$ (Property 4 of Proposition 0.5), that is, ω needs to be close to the k-th harmonic space of the simplicial complex for the above derivation to be valid for a full simplicial complex (and not just one simplex).

Summing up, the simplicial Kuramoto model may be obtained as a system of *harmonically resonant* ($L^k\omega \approx 0$) interacting oscillators on the effective hypergraph associated with the simplicial complex and an initially generic coupling function. It is interesting to note that, although the coupling functions obtained here are not phae shift-invariant as discussed above, they are also observed in some settings of node Kuramoto models, like in Ref. [114] for interacting communities of oscillators.

1.3.3 As above, *not* so below: the two types of interactions

The introduction of Definition 1.1 was initially motivated by purely formal and symmetry arguments. It is then important to study the *local* form of the different interaction terms to understand what kind of system is being described. Following a similar procedure to the one proposed in Ref. [83], we treat the two types of interactions separately.

From above. Let us start with the interaction from *above*

$$I_{\uparrow}(\theta) := -B^{k+1}\sin(D^k\theta)\,, \tag{1.11}$$

which is a direct generalization of the standard node Kuramoto interaction term $-B^1\sin(D^0\theta)$. To understand its behavior, we look at the simplest possible interaction of its kind, where we have a single $(k+1)$-simplex regulating the interaction between its $k+2$ oscillating faces

(Figure 1.2b). In this case, the incidence matrix is a column vector of the form

$$B_{k+1} = \xi^{\uparrow} \in \{-1, 1\}^{k+2} \,, \tag{1.12}$$

where $\xi_i^{\uparrow}$ is 1 if the face i is coherently oriented with the $(k+1)$-simplex, and -1 if it is incoherently oriented (Remark 0.1). It follows that $D^k = (B_{k+1})^\top = (\xi^{\uparrow})^\top$. Equation (1.11) then becomes

$$I_{\uparrow}(\theta) = -\xi^{\uparrow} \sin\left((\xi^{\uparrow})^\top \theta\right) \,, \tag{1.13}$$

which means that each oscillator will be influenced by the same scalar value given by the oriented sum of the phases $(\xi^{\uparrow})^\top \theta$, with a sign depending on the coherence or incoherence of the orientations (see Figure 1.2c). In the nodes case, this simply reduces to $I_{\uparrow}(\theta) = (-\sin(\theta_1 - \theta_2), -\sin(\theta_2 - \theta_1))^\top$. Given that a $(k+1)$-simplex always has $k+2$ faces, this kind of interaction involves $k+2$ oscillators and thus, for $k > 0$, is *genuinely higher-order*, in the sense that it does not result from the composition of multiple pairwise interaction terms.

From below. The interaction from *below*

$$I_{\downarrow}(\theta) := -D^{k-1} \sin(B^k \theta) \tag{1.14}$$

describes the interactions of simplicial oscillators through lower-order simplices. This interaction from below is absent in the node Kuramoto, and it represents the true novelty of the simplicial model. First, while only $k+2$ simplices of order k can interact through a $(k+1)$-simplex, an arbitrary number of k-simplices can have a common face and interact from below. This allows us to consider arbitrary higher-order interactions, not restricted by the order of the oscillating simplices.

It is then natural to ask if the interactions from below are locally of a similar form to the interactions from above, as in the case of Equation (1.11). For this, let us consider again the simplest possible interaction, i.e. the general case of N k-simplices lower adjacent through a common face with arbitrary orientations, as illustrated in Figure 1.2d. By considering an

appropriate ordering of the simplices, this configuration is described by the incidence matrix

$$B_k = \begin{pmatrix} \xi_1^{\downarrow} & \xi_2^{\downarrow} & \cdots & \xi_N^{\downarrow} \\ o_1 & 0 & \cdots & 0 \\ 0 & o_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & o_N \end{pmatrix}, \tag{1.15}$$

where the entries $\xi_i^{\downarrow} \in \{-1, 1\}$ describe the relative orientation between simplex i and the common face, while $o_i \in \{-1, 1\}^k$ contains the relative orientations between simplex i and its other faces not involved in this interaction from below. Given that $D^{k-1} = (B_k)^\top$, Equation (1.14) becomes

$$I_{\downarrow}(\theta) = -\xi^{\downarrow} \sin\left((\xi^{\downarrow})^\top \theta\right) - k \sin(\theta)\,, \tag{1.16}$$

where $\xi^{\downarrow} = (\xi_1^{\downarrow}, \ldots, \xi_N^{\downarrow})^\top$. Notice that the first term is formally the same as in Equation (1.13) for the interaction from above. Each oscillator receives a contribution depending on the phases of all oscillators involved in the interaction. This means that the higher-order interaction given by a $(k+1)$-simplex shares the same structure as the higher-order interaction given by $k+2$ oscillators sharing a common face.

Self interactions. However, an extra term $-k \sin(\theta)$ appears. By carrying out the computations that lead from Equation (1.14) to Equation (1.16), it appears that this extra term is a sum of contributions coming from the faces not involved in the interaction, which, in this case, are **free** i.e. they are faces of only one simplex. In fact, here each oscillator has k free faces, hence the multiplication factor k in front of $\sin(\theta)$. In general, due to this term, each oscillator modulates its frequency based on its own phase, which is akin to a *self-interaction* through its free faces (Figure 1.2e). Formally, this term also appears in the Adler equation [15] describing the phase difference of a system with one oscillator driven by another one. From that point of view, the self-interaction terms can be seen as the driving of each oscillator by another non-existent oscillator that has a constant phase set to zero.

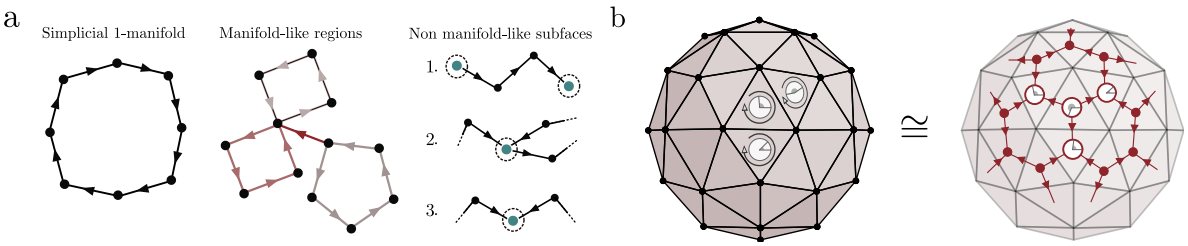

Fig. 1.3 The form of the interactions from below of the simplicial Kuramoto is equivalent to the ones of the standard node Kuramoto on manifold-like faces of the simplicial complex. **a.** From left to right: a 1-dimensional simplicial manifold, where every node is manifold-like as it is incident to exactly two edges with different orientations. In the middle is a simplicial complex where the 1-dimensional manifold-like regions are highlighted with different colors. On the right, the different ways in which a face can produce an interaction different from a standard Kuramoto interaction: 1. the face is free, 2. there are more than two oscillators incident to it, 3. there are two oscillators incident to it which are both coherently or incoherently oriented. **b.** If the complex is an oriented simplicial manifold, then the interaction term from below is equivalent to a node Kuramoto taking place on the 1-skeleton of the dual cell complex.

1.3.4 Manifold-like simplicial complexes

Interestingly, if the interaction from below involves exactly two oscillating simplices, one coherent and the other incoherent with respect to the common face, the interaction term will be the same pairwise one as the standard node Kuramoto, i.e. of the form $\sin(\theta_1 - \theta_2)$. Formally, this situation coincides with the simplicial complex being locally Euclidean, akin to a manifold.

Definition 1.2 (Simplicial manifold)**.**

An oriented simplicial complex X is said to be a (oriented) **simplicial manifold** of dimension d if any $(d-1)$-simplex in X is the face of exactly two d-simplices, one coherent and one incoherently oriented with it. In this case, we say that such a face is **manifold-like**.

Different kinds of interactions occur at non-manifold-like faces (Figure 1.3a), that is:

1. at faces that are free, resulting in *self-interactions*;
2. at faces that are adjacent to more than two simplices of order k, i.e. *genuinely high-order interactions*;
3. at faces adjacent to two simplices that are both coherently or incoherently oriented, resulting in interactions of the form $\sin(\theta_1 + \theta_2)$ [115]. Remembering that orientations

are just reference frames, these correspond to standard Kuramoto interactions when the phase of one oscillator is measured in the "reverse" sense.

If every $(k-1)$-simplex which has at least a k-simplex incident to it is manifold-like, so that the simplicial complex is a simplicial manifold, we have the following equivalence result.

Theorem 1.1 (Simplicial Kuramoto on a manifold)**.** *Let X be a k-dimensional oriented simplicial manifold. Then it follows that the simplicial Kuramoto dynamics of order k is equivalent to the standard node Kuramoto taking place on the* 1*-skeleton of the dual cell complex to X, that is, the graph with a node for each k-simplex and an edge for each $(k-1)$-simplex.*

Proof. Under the assumptions that X is manifold-like, we can apply the discrete analogous to Poincaré duality (see [2] p.50) to obtain $B^k = \hat{D}^0, \quad D^{k-1} = \hat{B}^1$, where $\hat{B}$ and $\hat{D}$ are, respectively, the boundary and coboundary operators of the dual cell complex to X. The interaction term from below in the primal complex becomes the interaction term from above in the 1-skeleton of the dual, i.e.

$$\dot{\theta} = \omega - \sigma^{\downarrow} D^{k-1} \sin(B^k \theta) = \omega - \sigma^{\downarrow} \hat{B}^1 \sin(\hat{D}^0 \theta)\,,$$

which has the same form as the standard node Kuramoto model in Equation (1.4). □

An illustration of this result can be seen in Figure 1.3b with a triangulated sphere. Notice also that the dual graph to a simplicial k-manifold will necessarily be a $(k+1)$-regular graph, as every oscillating k-simplex has exactly $k+1$ faces.

1.3.5 Hodge decomposition of the dynamics

Thanks to the particular form of the two interaction terms, one can use a well-known result in combinatorial topology to decompose the dynamics into three independent subdynamics.

To show this, let us consider a simplicial complex X, weighted or unweighted, which describes the interactions between k-th order oscillators. Then, we can apply the simplicial Hodge decomposition theorem we stated above (Theorem 0.3) to decompose the signal space into divergence-free, harmonic, and curl-free orthogonal components. We use the theorem to decompose both the phases cochain θ and the natural frequencies ω

$$\theta = \theta_{\mathrm{df}} + \theta_{\mathrm{H}} + \theta_{\mathrm{cf}},\ \omega = \omega_{\mathrm{df}} + \omega_{\mathrm{H}} + \omega_{\mathrm{cf}}\,, \tag{1.17}$$

where "cf" stands for *curl-free*, "H" for *harmonic*, and "df" for *divergence-free*. Rewriting the simplicial Kuramoto dynamics leveraging the orthogonality of the components, Definition 1.1 is equivalent to the following system of differential equations:

$$\begin{cases} \dot{\theta}_{\mathrm{df}} = \omega_{\mathrm{df}} - \sigma^{\uparrow} B^{k+1} \sin(D^k \theta_{\mathrm{df}}) \\ \dot{\theta}_{\mathrm{H}} = \omega_{\mathrm{H}} \\ \dot{\theta}_{\mathrm{cf}} = \omega_{\mathrm{cf}} - \sigma^{\downarrow} D^{k-1} \sin(B^k \theta_{\mathrm{cf}}) \end{cases} . \tag{1.18}$$

These three equations are of crucial importance. They tell us that under the simplicial Kuramoto dynamics: **i)** the curl-free, the harmonic, and the divergence-free components evolve independently of one another, and **ii)** the harmonic component is not affected by the interaction terms. Notice also that the interaction from above affects only the divergence-free component, while the one from below affects only the curl-free component.

Harmonic gauge of the model. If $\omega_{\mathrm{H}} \neq 0$, there can be no equilibrium of the system, as each component of θ_{H} will always evolve with a fixed angular speed. It follows that it is always possible to pass to a frame of reference where the harmonic component evolves with constant speed in time, simply by performing the change of variables $\theta \to \theta - \omega_{\mathrm{H}} t$. In the case of the node Kuramoto (Equation (1.3)), this corresponds to passing to what is known as the **co-rotating frame of reference**, as $\omega_{\mathrm{H}} = \bar{\omega}\mathbb{1}$, i.e. the constant vector of the average natural frequency. When studying the node Kuramoto dynamics, this means that one can focus w.l.o.g. on the case where the average natural frequency is zero.

This is part of a more general observation that the addition of a harmonic cochain $x \in \ker L^k$ to the phases has no effect on the dynamics. In fact, property 4 of Proposition 0.5 tells us that $\ker L^k = \ker B^k \cap \ker D^k$ and thus both $B^k x$ and $D^k x$ are zero. Any change of variable $\gamma = \theta + x$ will thus leave the dynamics (Definition 1.1) formally unchanged. In this sense, we can say that the harmonic space is the natural *gauge* of the simplicial Kuramoto.

Remark 1.1 (Phase-shift symmetry)**.** *The fact that in general* $\mathbb{1}$ *is not harmonic in higher-order systems on simplicial complexes means that these systems do not possess a* phase-shift *symmetry in the traditional sense: shifting all oscillators by the same phase (summing a vector proportional to* $\mathbb{1}$*) does* ***not*** *leave the dynamics unchanged, contrary to traditional node Kuramoto models and immediate physical intuition.*

However, a (possibly bigger) symmetry space exists with the harmonic space, which gives a different non-constant "harmonic" phase-shift symmetry to the model for each hole present in the simplicial complex.

Note that the presence of additional symmetries may also occur when considering specific higher-order interactions between populations of node-based oscillators [43, 44].

1.3.6 Simplicial order parameters and gradient system

To measure the degree of synchronization of a phase configuration, it is common to employ the *order parameter*, which, for an unweighted network of N oscillators, is defined as

$$\widetilde{R}(\theta) = \frac{1}{N}\left|\sum_{\alpha=1}^{N} e^{i\theta_\alpha}\right| . \tag{1.19}$$

By definition, it is non-negative, and it reaches its maximum value of 1 when the oscillators are *fully synchronized*, i.e. when they all have the same phase $\theta \propto \mathbb{1}$.

The order parameter defined in Equation (1.19), however, assumes an all-to-all coupling of the oscillators and thus is not appropriate to describe situations where the topology of the interactions is non-trivial. In this chapter, we focus on a particular generalization [76] of Equation (1.19) which takes into account the underlying network structure through the use of the incidence matrix

$$R(\theta) = \frac{n_0^2 - 2n_1 + 2\mathbb{1}^\top \cos(B_1^\top \theta)}{n_0^2} . \tag{1.20}$$

While Equation (1.20) reduces to Equation (1.19) in the case of a fully connected network, it is, in fact, a natural order parameter for Equation (1.4), as its gradient gives us the interaction term of the dynamics

$$\nabla_\theta R(\theta) \propto -B_1 \sin(B_1^\top \theta) = I_\uparrow(\theta) , \tag{1.21}$$

meaning that $R(\theta)$ is the potential function which gives the node Kuramoto interaction term as its gradient, i.e. $\dot{\theta} = \omega + \sigma \frac{n_0^2}{2} \nabla_\theta R(\theta)$. When we can write the interaction term of a model as a gradient of a potential function, we say, with a slight abuse of terminology, that its dynamics is in "gradient system" form. In these cases, we interpret the potential as the natural order parameter associated with the model.

As proposed in Ref. [50], we can extend this intuition to the simplicial case and, neglecting constants that do not appear in the gradient, define the *simplicial order parameter*

$$R_k(\theta) = \frac{1}{C_k}\left(\mathbb{1}^\top W_{k-1}^{-1}\cos(B^k\theta) + \mathbb{1}^\top W_{k+1}^{-1}\cos(D^k\theta)\right), \tag{1.22}$$

with the normalization constant

$$C_k = \mathbb{1}^\top W_{k-1}^{-1}\mathbb{1} + \mathbb{1}^\top W_{k+1}^{-1}\mathbb{1}\,. \tag{1.23}$$

The weight matrices are added to further generalize the construction to weighted simplicial complexes, and generate the weighted simplicial Kuramoto model as the gradient system

$$W_k\nabla_\theta R_k(\theta) \propto I_\uparrow(\theta) + I_\downarrow(\theta)\,. \tag{1.24}$$

This order parameter reaches a maximum value of 1 if $\theta \in \ker B^k \cap \ker D^k = \ker L^k$ i.e. when the phases cochain belongs to the harmonic space. This is a direct generalization of synchronization in the node Kuramoto model: in a connected network, $\ker L^0 = \text{span}\{\mathbb{1}\}$, and the full synchronization condition is $\theta \propto \mathbb{1}$, which is equivalent to the phase cochains being harmonic.

Hence, under this definition, full synchronization in the simplicial model does not mean that the phases are all equal, but that θ is harmonic [50]. Moreover, as the k-th harmonic space of a simplicial complex is isomorphic to the k-th homology group (Theorem 0.4), we can think of fully synchronized configurations as, intuitively, localized around the k-dimensional holes.

Definition 1.3 (Full synchronization)**.**

A configuration θ is said to be *fully synchronized* under the k-th order simplicial Kuramoto dynamics if $\theta \in \ker L^k$.

From the simplicial order parameter Equation (1.22), we can extract two *partial order parameters*

$$R_k^-(\theta) := \frac{1}{C_k^-}\mathbb{1}^\top W_{k-1}^{-1}\cos\left(B^k\theta\right) \tag{1.25a}$$

$$R_k^+(\theta) := \frac{1}{C_k^+}\mathbb{1}^\top W_{k+1}^{-1}\cos\left(D^k\theta\right), \tag{1.25b}$$

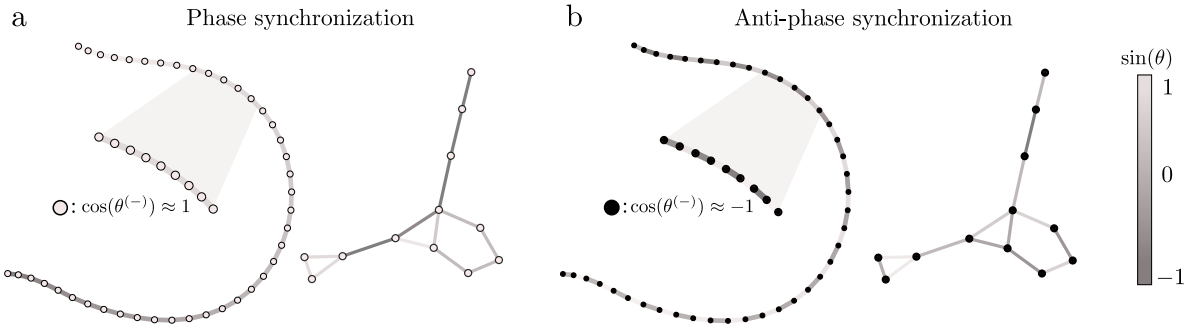

Fig. 1.4 Configurations of phases $\theta \in C^1$, whose sine is shown here in color, which are phase ($R_1(\theta) \approx 1$) and anti-phase ($R_1(\theta) \approx -1$) synchronized in the case of a chain of edges, which is "quasi"-manifold as all faces except the endpoints are manifold-like, and a more general 1-dimensional simplicial complex. **a.** In the case of phase synchronization, close oscillators on manifold-like regions have similar phases. **b.** Anti-phase synchronized configurations, instead, are such that, on manifold-like regions, adjacent oscillators have opposite phases.

where the normalization constants $C_k^{\pm} = \mathbb{1}^\top W_{k\pm1}^{-1} \mathbb{1}$ ensure that they take values in $[-1, 1]$. In this way, it holds that

$$C_k R_k(\theta) = C_k^+ R_k^+(\theta) + C_k^- R_k^-(\theta)\,, \tag{1.26}$$

and thus, aside from normalization, the order of a configuration is computed by measuring separately the local order induced respectively on $(k-1)$ and $(k+1)$-simplices.

Notice that, by neglecting the constants in passing from Equation (1.20) to Equation (1.22), we have an order parameter that has values in the interval $[-1, 1]$. This allows us to meaningfully distinguish two different types of synchronized configurations. We call a configuration of phases *phase synchronized* when its order is close to 1 and *anti-phase synchronized* when it is close to -1. Phase synchronization generalizes to simplicial complexes the situation where close oscillators have similar phases (Figure 1.4a), while in anti-phase synchronization close oscillators have opposite phases forming "checkerboard" patterns, resembling an antiferromagnetic Ising model (Figure 1.4b).

Notice also how, in this chapter, with "phase synchronization" and "full synchronization" we refer to the static properties of a configuration of phases θ, with no information on how it evolves under the dynamics. The notion of a configuration that "stays synchronized" under the dynamics will be tackled with the concept of phase-locking in Section 1.4.1.

1.4 Equilibrium analysis

We now study the equilibrium properties of the simplicial Kuramoto model, extending to the simplicial cases concepts and results known in the node case.

- In Section 1.4.1, we extend the notion of phase-locking to simplicial complexes, we look at its geometric meaning and see how it reduces to standard node synchronization on manifold-like regions of the complex.
- In Section 1.4.2, we develop the necessary framework to discuss the equilibrium properties of the simplicial Kuramoto model, define reachable equilibria (Definition 1.8) and relate their existence to the presence of simplicial phase-locked configurations.
- In Section 1.4.3, we derive two bounds on the coupling strength, providing necessary conditions for the existence of equilibria. We define the critical coupling (Definition 1.9) and characterize it as the solution to a linear optimization problem.
- In Section 1.4.4, we prove a simple lower bound on the coupling strength which gives a sufficient condition for the existence of reachable equilibria.

1.4.1 Simplicial phase-locking

It directly follows from the Hodge decomposition of the simplicial Kuramoto model of Equation (1.18) that studying its equilibrium properties is equivalent to separately studying the equilibria of the curl-free and divergence-free components. If these two converge to equilibrium, then the complete system will converge to a configuration evolving with constant harmonic angular speed, given by ω_H ($\dot{\theta} = \dot{\theta}_H = \omega_H$).

Definition 1.4 (Simplicial phase-locking)**.**

We say that the k-th order simplicial Kuramoto dynamics is *phase-locked from above* if $\dot{\theta}_{df} = 0$ and *phase-locked from below* if $\dot{\theta}_{cf} = 0$.

To motivate this definition, we consider the projections of the dynamics on lower and upper order simplices, defined as

$$\theta^{(+)} := D^k\theta = D^k\theta_{\mathrm{df}} \tag{1.27}$$

$$\theta^{(-)} := B^k\theta = B^k\theta_{\mathrm{cf}}\,. \tag{1.28}$$

We can think of $\theta^{(+)}$ and $\theta^{(-)}$ as the discrete versions of, respectively, the curl and divergence of the vector field θ and, we prove, they can equivalently capture simplicial phase-locking.

Proposition 1.2 (Phase-locking equivalence)**.** *A configuration of phases θ is phase-locked from above (from below) if and only if its projection onto higher (lower) dimensional simplices is in equilibrium.*

$$\dot\theta_{\mathrm{df}} = 0 \iff \dot\theta^{(+)} = 0, \quad \dot\theta_{\mathrm{cf}} \iff \dot\theta^{(-)} = 0\,. \tag{1.29}$$

Proof. If $\dot\theta_{\mathrm{df}} = 0$, then

$$\dot\theta^{(+)} = D^k\dot\theta_{\mathrm{df}} = 0\,,$$

because of Equation (1.27). If instead $\dot\theta^{(+)} = 0$, then

$$\dot\theta_{\mathrm{df}} = (D^k)^\dagger D^k\dot\theta = (D^k)^\dagger\dot\theta^{(+)} = 0\,,$$

where $(D^k)^\dagger$ is the weighted Moore-Penrose pseudoinverse [116] and $(D^k)^\dagger D^k$ is the orthogonal projection operator onto $\mathrm{Im}(D^k)^* = \mathrm{Im}\, B^{k+1}$. □

This result allows us to include in Definition 1.4 the standard concept of phase-locking for the node Kuramoto. In fact, the node Kuramoto on a heterogeneous network is classically said to be phase-locked when the phase difference of connected oscillators stays constant in time. This means that $\dot\theta_e^{(+)} = (D^0\dot\theta)_e = \dot\theta_i - \dot\theta_j = 0$ for every edge $e = (i, j)$ in the network. According to Proposition 1.2, the divergence-free component of the dynamics is in equilibrium and the system is, by Definition 1.4, *simplicially* phase-locked from above. If the network is connected, moreover, $\dot\theta$ must be harmonic, i.e. $\dot\theta \propto \mathbb{1}$, a situation which is usually named *frequency-synchronized* as the frequencies of all oscillators coincide.

While phase-locking in the standard Kuramoto model means that all oscillators evolve with the same angular frequency, it is not clear how this extends to the simplicial case. In the case

of phase-locking from below, we see that

$$\dot{\theta}^{(-)} = 0 \iff \frac{d}{dt}(B^k\theta) = B^k\dot{\theta} = 0\,,$$

or equivalently that $\dot{\theta}(t)$, the cochain containing the angular frequencies, is a cycle. This, in turn, means that for each $(k-1)$-simplex α

$$\sum_{i>\alpha} \xi_{\alpha,i}\dot{\theta}_i = 0\,, \tag{1.30}$$

where $\xi_{\alpha,i} \in \{-1, 1\}$ is the relative orientation of k-simplex i with respect to its face α. Interestingly, Equation (1.30) corresponds to a flow conservation condition. Indeed, if we consider graphs with oscillating edges, at each node, the total phase flow of the incoming edges is the same as the total flow of the outgoing ones. In general, we can apply Equation (1.30) to understand phase-locking from below in some particular situations.

Proposition 1.3 (Phase-locking on manifold-like regions)**.** *If θ is phase-locked from below, i.e. $\dot{\theta}^{(-)} = 0$,*

1. *the connected manifold-like regions of the complex (with respect to order k) evolve with the same angular frequency and are thus frequency-synchronized;*
2. *oscillators with free faces are frozen, i.e. $\dot{\theta}_i = 0$.*

Proof. 1. At a manifold-like face we have only two incident simplices, one coherently oriented and one incoherent, which we name respectively i and j. Condition Equation (1.30) gives us $\dot{\theta}_i - \dot{\theta}_j = 0 \iff \dot{\theta}_i = \dot{\theta}_j$, so the incident oscillators are frequency-synchronized. 2. If oscillator i has a free face then, by definition, that particular face will be incident only to oscillator i. Phase-locking at that face implies that $\dot{\theta}_i = 0$, thus concluding the proof. □

A simple application of this result is shown in Figure 1.5a where the behavior of a phase-locked configuration can be inferred a priori by looking at the geometry of the graph. It can also be empirically seen that frequency-synchronized manifold regions exhibit phenomena akin to traveling waves localized around the k-holes of the complex, when its homology is not trivial.

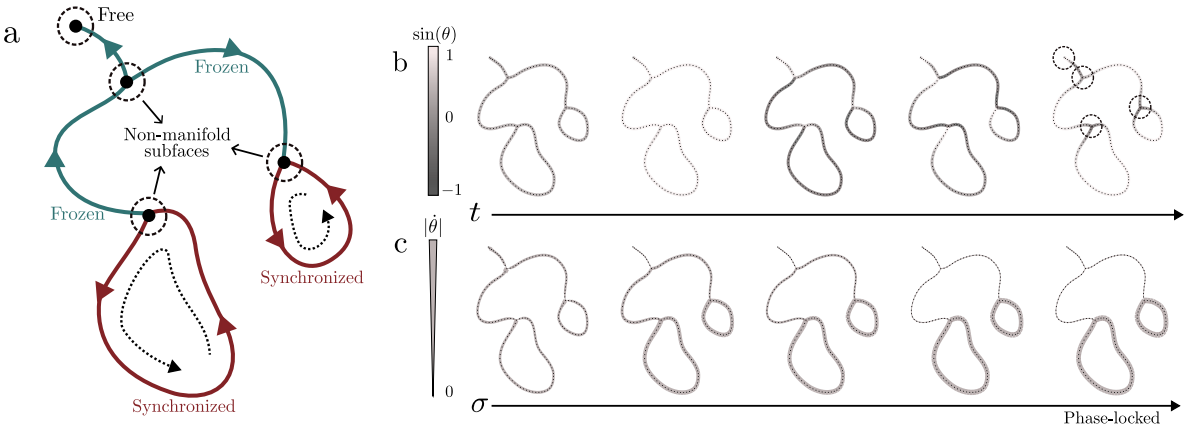

Fig. 1.5 Simplicial Kuramoto dynamics on a simple graph with two holes (drawn here as a continuous space), where the edges are identical oscillators ($\omega = \mathbb{1}$) with starting phase $\theta(0) = 0\mathbb{1}$. **a.** The panel shows a diagram of the graph, highlighting the non-manifold points responsible for the non-triviality of the dynamics. The different branches of the graph are colored according to their frequency in a phase-locked (Definition 1.4) state. In particular, the oscillation is frequency-synchronized on the two holes and exhibits traveling waves, while it is frozen ($\dot{\theta} = 0$) on the branch connecting them. **b.** A few snapshots of the dynamics on the graph are shown, with edges colored according to the sine of their phases. In the last frame, the effect of the non-manifold points is evident. **c.** The dynamics is run for different values of the coupling strength, and the absolute value of the frequency ($|\dot{\theta}|$) at the final integration time is shown with the edges' widths. The last frame shows how the system reaches the same phase-locked configuration predicted with Equation (1.30) and depicted in panel a.

1.4.2 Existence of equilibria

To study the equilibrium of the simplicial Kuramoto model, it is convenient to work with the $\theta^{(\pm)}$, defined in Equations (1.27) and (1.28) as *projections* of the phases onto upper and lower simplices. Their evolution equations [33] are readily obtained by multiplying the dynamics (Definition 1.1) by D^k and B^k to get

$$\begin{aligned} \dot{\theta}^{(+)} &= \omega^{(+)} - \sigma^{\uparrow} L_{\downarrow}^{k+1} \sin(\theta^{(+)}) := f^{(+)}(\theta^{(+)}) \\ \dot{\theta}^{(-)} &= \omega^{(-)} - \sigma^{\downarrow} L_{\uparrow}^{k-1} \sin(\theta^{(-)}) := f^{(-)}(\theta^{(-)}) \,, \end{aligned} \tag{1.31}$$

where $L_{\downarrow}^{k+1}, L_{\uparrow}^{k-1}$ are the half Laplacian matrices from Definition 0.19, and we defined the projected natural frequencies as

$$\omega^{(+)} := D^k \omega, \qquad \omega^{(-)} := B^k \omega \,. \tag{1.32}$$

An equilibrium for the projected dynamics, say, $\theta_*^{(+)} \in \mathbb{R}^{n_{k+1}}$, will thus need to satisfy the equation $f^{(+)}(\theta_*^{(+)}) = 0$, while also being a projection i.e. $\theta_*^{(+)} \in \operatorname{Im} D^k$. Let us now focus on the first condition, and compute all the possible solutions to the equations $f^{(\pm)}(\theta_*^{(\pm)}) = 0$. We have

$$\begin{cases} f^{(+)}(\theta_*^{(+)}) = 0 \iff L_\downarrow^{k+1} \sin(\theta_*^{(+)}) = \frac{\omega^{(+)}}{\sigma^\uparrow} \\ f^{(-)}(\theta_*^{(-)}) = 0 \iff L_\uparrow^{k-1} \sin(\theta_*^{(-)}) = \frac{\omega^{(-)}}{\sigma^\downarrow}\,, \end{cases} \tag{1.33}$$

and, since $\omega^{(+)} \in \operatorname{Im} D^k = \operatorname{Im} L_\downarrow^{k+1}$ and $\omega^{(-)} \in \operatorname{Im} B^k = \operatorname{Im} L_\uparrow^{k-1}$, the equilibrium equations can be solved using the pseudoinverse

$$\begin{cases} \sin(\theta_*^{(+)}) = (L_\downarrow^{k+1})^\dagger \frac{\omega^{(+)}}{\sigma^\uparrow} + x^{(+)} \\ \sin(\theta_*^{(-)}) = (L_\uparrow^{k+1})^\dagger \frac{\omega^{(-)}}{\sigma^\downarrow} + x^{(-)} \end{cases}, \tag{1.34}$$

for any weighted $(k+1)$-cycle $x^{(+)} \in \ker B^{k+1}$ and $(k-1)$-cocycle $x^{(-)} \in \ker D^{k-1}$. Moreover, by applying well-known properties of the Moore-Penrose pseudoinverse, we can simplify these expressions.

Lemma 1.1. *We have the following equalities*

$$(L_\downarrow^{k+1})^\dagger \omega^{(+)} = (B^{k+1})^\dagger \omega, \quad (L_\uparrow^{k-1})^\dagger \omega^{(-)} = (D^{k-1})^\dagger \omega\,. \tag{1.35}$$

Proof. We have

$$\begin{aligned} (L_\downarrow^{k+1})^\dagger \omega^{(+)} &= (D^k B^{k+1})^\dagger D^k \omega = (B^{k+1})^\dagger (D^k)^\dagger D^k \omega \\ &= (D^k)^{*\dagger} (D^k)^\dagger D^k \omega = (B^{k+1})^\dagger \omega, \end{aligned}$$

and, analogously, $(L_\uparrow^{k-1})^\dagger \omega^{(-)} = (D^{k-1})^\dagger \omega$. □

Definition 1.5 (Natural potentials)**.**

We call *natural potentials* of order k the quantities

$$\beta^{(+)} = (B^{k+1})^\dagger \omega \in \mathbb{R}^{n_{k+1}},\ \beta^{(-)} = (D^{k-1})^\dagger \omega \in \mathbb{R}^{n_{k-1}}\,. \tag{1.36}$$

The name *potential* comes from the fact that, for $k = 1$, $\beta^{(-)}$ is an assignment of potentials to the nodes such that for each edge the difference of potential between its end-points (i.e., the

voltage) is equal to ω. Moreover, it can be easily proven that they correspond to the higher- and lower-order signals that appear in the Hodge components of the natural frequency vector ω as

$$\omega = B^{k+1}\beta^{(+)} + \omega_{\mathrm{H}} + D^{k-1}\beta^{(-)} .$$

The values of the natural potentials are expressed in terms of the weighted Moore-Penrose pseudoinverse, which in Equation (1.36) is computed with respect to the inner products on the cochain spaces W_k^{-1} $k = 1, \dots, K$ (Definition 0.17). To compute the natural potentials of a weighted simplicial complex, the weights have to be included correctly in the pseudoinverse. The explicit formula, written in terms of the standard unweighted pseudoinverse is the following (Ref. [116, Remark 2])

$$\beta^{(+)} = W_{k+1}^{\frac{1}{2}} \left(W_k^{-\frac{1}{2}} B^{k+1} W_{k+1}^{\frac{1}{2}} \right)^{\dagger} W_k^{-\frac{1}{2}} \omega \tag{1.37}$$

$$\beta^{(-)} = W_{k-1}^{\frac{1}{2}} \left(W_k^{-\frac{1}{2}} D^{k-1} W_{k-1}^{\frac{1}{2}} \right)^{\dagger} W_k^{-\frac{1}{2}} \omega . \tag{1.38}$$

Admissibility. Using the definition of natural potentials, we can rewrite the equilibrium conditions of Equation (1.34) as

$$\begin{cases} \sin(\theta_*^{(+)}) = \frac{\beta^{(+)}}{\sigma^{\uparrow}} + x^{(+)} \\ \sin(\theta_*^{(-)}) = \frac{\beta^{(-)}}{\sigma^{\downarrow}} + x^{(-)} \end{cases} , \tag{1.39}$$

where we see that a *necessary* condition for the existence of a solution is for the right-hand sides to be element-wise bounded in $[-1, 1]$ or, equivalently, for $x^{(+)} \in \ker B^{k+1}$ and $x^{(-)} \in \ker D^{k-1}$ to satisfy the following condition of *admissibility*.

Definition 1.6 (Admissible cycles)**.**

We call a cycle $x^{(+)} \in \ker B^{k+1}$ *admissible* if

$$\left\| \frac{\beta^{(+)}}{\sigma^{\uparrow}} + x^{(+)} \right\|_{\infty} \leq 1 \,. \tag{1.40}$$

We call a cocycle $x^{(-)} \in \ker D^{k-1}$ *admissible* if

$$\left\| \frac{\beta^{(-)}}{\sigma^{\downarrow}} + x^{(-)} \right\|_{\infty} \leq 1 \,. \tag{1.41}$$

With a slight abuse of notation, we call them both *admissible cycles*, and we name their sets $\mathcal{A}^{(+)}$ and $\mathcal{A}^{(-)}$.

Proposition 1.4 (Necessary condition from admissible cycles)**.** *A* necessary *condition for the existence of equilibrium solutions of the* $(\pm)$ *dynamics is that* $\mathcal{A}^{(\pm)} \neq \emptyset$.

Intuitively, each cochain $\beta^{(\pm)}/\sigma^{\updownarrow}$ should be *close* to, respectively, the vector space of weighted $(k+1)$-cycles (for $(+)$) and $(k-1)$-cocycles (for $(-)$).

When both $x^{(+)}$ and $x^{(-)}$ are admissible, we can invert the sine function in Equation (1.34) and get an explicit expression for the set of equilibrium configurations of the projections.

Definition 1.7 (Equilibrium sets)**.**

We define the equilibrium sets of the projections as

$$\mathcal{E}^{(+)} = \left\{ (-1)^{s_+} \odot \arcsin\left(\frac{\beta^{(+)}}{\sigma^{\uparrow}} + x^{(+)} \right) + \pi s_+ + 2\pi m_+ : \right. \\ \left. s_+ \in \{0,1\}^{n_{k+1}} \,,\; m_+ \in \mathbb{Z}^{n_{k+1}},\; x^{(+)} \in \mathcal{A}^{(+)} \right\} \,, \tag{1.42}$$

$$\mathcal{E}^{(-)} = \left\{ (-1)^{s_-} \odot \arcsin\left(\frac{\beta^{(-)}}{\sigma^{\downarrow}} + x^{(-)} \right) + \pi s_- + 2\pi m_- : \right. \\ \left. s_- \in \{0,1\}^{n_{k-1}} \,,\; m_- \in \mathbb{Z}^{n_{k-1}},\; x^{(-)} \in \mathcal{A}^{(-)} \right\} , \tag{1.43}$$

where $\odot$ is the component-wise Hadamard product.

Any $\theta_*^{(\pm)} \in \mathcal{E}^{(\pm)}$ will thus be a fixed point of the dynamics of Equation (1.31).

From a geometric point of view, the equilibrium set $\mathcal{E}^{(\pm)}$ is a subset of $\mathbb{R}^{n_{k\pm1}}$ and, for fixed $s_\pm, m_\pm$, is a manifold of dimension given by $\dim\ker D^{k-1}$ for $(-)$ and $\dim\ker B^{k+1}$ for $(+)$. For example, for the projection on the nodes of edge dynamics, we have that $\dim\ker D^0$ is the number of connected components. If the simplicial complex is connected then $\mathcal{E}^{(-)}$ is a collection of curves in an n_0-dimensional space (see Figure 1.6b).

We thus found a complete characterization of the configurations $\theta_*^{(\pm)} \in \mathbb{R}^{n_{k\pm1}}$ which are fixed points of the dynamics of Equation (1.31).

Proposition 1.5. *$f^{(\pm)}(\theta_*^{(\pm)}) = 0$ if and only if $\theta_*^{(\pm)} \in \mathcal{E}^{(\pm)}$.*

Reachability. This condition is only partially useful because, while it tells us which configurations $\theta_*^{(\pm)}$ are solutions to the equilibrium equations, it does not specify which ones are projections, i.e. can be obtained by projecting a phase cochain θ on the $(k \pm 1)$-simplices. In particular, notice that the dynamics for the $(-)$ component (the same holds for $(+)$) in Equation (1.31) states that the time derivative of $\theta^{(-)}$ will be the vector $\omega^{(-)} - \sigma^{\downarrow} L_{\uparrow}^{k-1} \sin(\theta^{(-)})$, which always belongs to $\operatorname{Im} B^k$. This, together with the initial phase configuration being a projection of a phase cochain, $\theta_0^{(-)} = B^k\theta_0 \in \operatorname{Im} B^k$, implies that the trajectories live in the subspace $\operatorname{Im} B^k$ i.e. projections remain projections under the dynamics. Only the equilibria in $\mathcal{E}^{(-)}$ which also belong to $\operatorname{Im} B^k$ are thus relevant for the dynamics, and thus we call them *reachable*.

Definition 1.8 (Reachable equilibria)**.**

We define the sets of reachable equilibria as

$$\mathcal{R}^{(-)} = \mathcal{E}^{(-)} \cap \operatorname{Im} B^k \tag{1.44}$$

$$\mathcal{R}^{(+)} = \mathcal{E}^{(+)} \cap \operatorname{Im} D^k \,. \tag{1.45}$$

We then have our final result.

Proposition 1.6 (Equivalence phase-locking reachability)**.** *The curl-free (divergence-free) component admits equilibria if and only if $\theta^{(-)}$ (resp. $\theta^{(+)}$) admits reachable equilibria.*

The framework we developed in this section can be fruitfully exploited to independently discuss the existence of equilibrium configurations of the divergence-free and of the curl-free components of the simplicial Kuramoto dynamics (Equation (1.18)). In particular, Proposition 1.6 tells us that such configurations will exist if and only if there are reachable

equilibria for the projections. These, in turn, are subsets of the larger sets of fixed points (Definition 1.7), whose explicit expression is known and whose non-emptiness can thus be controlled more easily, giving us necessary conditions for equilibrium.

Interestingly, one can prove that Definition 1.8 and Proposition 1.6 are simplicial generalizations of the *dynamic* and *geometric* conditions for equilibrium in the node Kuramoto that can be found in the literature (Ref. [117]), see Appendix A of Ref. [54].

1.4.3 Necessary conditions for phase-locking

In this section, we investigate the relation between the equilibrium properties of the simplicial Kuramoto model and the value of the coupling strength. It is natural to think that having a stronger interaction would make it easier for the system to reach a synchronized configuration as the intrinsic differences among the oscillators, encoded by their natural frequencies, become secondary. This intuition is extensively confirmed by numerous results proved about the node Kuramoto (see Refs. [118, 76, 119]), some of which we extend to the simplicial case. We will thus derive bounds on the coupling strength, which gives us necessary and sufficient conditions for the existence of reachable equilibria, i.e. for phase-locking from below and from above. Note that all the results below refer to the *existence* of phase-locked configurations and provide no information about whether the dynamics will actually converge to them.

Let us consider a weighted simplicial complex whose m-simplices have weights $w^m_1, \dots, w^m_{n_m}$ for any order m, and focus on the k-th order simplicial Kuramoto dynamics (Definition 1.1).

Simple bound. The easiest conditions to derive are those that ensure that there are no admissible cycles $\mathcal{A}^{(\pm)} = \emptyset$ (Definition 1.6). If this holds, then the equilibrium sets are empty $\mathcal{E}^{(\pm)} = \emptyset$ (Definition 1.7) and, by inclusion, the reachable sets are as well $\mathcal{R}^{(\pm)} = \emptyset$ (Definition 1.8) i.e. there are no reachable equilibria/phase-locked configurations.

Proposition 1.7 (Sufficient condition for no phase-locking)**.** *If*

$$\sigma^{\updownarrow} < \sigma_s^{(\pm)} := \frac{1}{\sqrt{n^w_{k\pm1}}} \left\| \beta^{(\pm)} \right\|_{(k\pm1)} , \tag{1.46}$$

where $n_{k\pm1}$ is the "weighted" number of simplices

$$n^w_{k\pm1} := \sum_{i=1}^{n_{k\pm1}} \frac{1}{w_i^{k\pm1}} , \tag{1.47}$$

then $\mathcal{E}^{(\pm)} = \emptyset$ *and the* $(\pm)$ *projection admits no equilibria.*

Proof. First, see that we can bound the weighted $(k \pm 1)$ norm (Definition 0.17) with the ∞-norm:

$$\|v\|_{(k\pm1)} = \sqrt{\sum_{i=1}^{n_{k\pm1}} \frac{1}{w_i^{k\pm1}} v_i^2} \leq \sqrt{n_{k\pm1}^w \left(\max_i v_i^2\right)} \leq \sqrt{n_{k\pm1}^w}\, \|v\|_\infty \,.$$

With this in mind, we can write

$$\left\| \frac{\beta^{(\pm)}}{\sigma^\updownarrow} + x^{(\pm)} \right\|_\infty \geq \frac{1}{\sqrt{n_{k\pm1}^w}} \left\| \frac{\beta^{(\pm)}}{\sigma^\updownarrow} + x^{(\pm)} \right\|_{(k\pm1)} .$$

The two addenda in the norm are orthogonal with respect to the inner product $\langle \cdot, \cdot \rangle_{(k\pm1)}$ because, in the $(-)$ case, $x^{(-)} \in \ker D^{k-1}$ and

$$\beta^{(-)} \in \operatorname{Im}(D^{k-1})^\dagger = \operatorname{Im}(D^{k-1})^* = (\ker D^{k-1})^\perp \,,$$

thus

$$\frac{1}{\sqrt{n_{k\pm1}^w}} \left\| \frac{\beta^{(\pm)}}{\sigma^\updownarrow} + x^{(\pm)} \right\|_{(k\pm1)} = \frac{1}{\sqrt{n_{k\pm1}^w}} \sqrt{\left\| \frac{\beta^{(\pm)}}{\sigma^\updownarrow} \right\|_{(k\pm1)}^2 + \left\| x^{(\pm)} \right\|_{(k\pm1)}^2} \geq \frac{1}{\sqrt{n_{k\pm1}^w}} \left\| \frac{\beta^{(\pm)}}{\sigma^\updownarrow} \right\|_{(k\pm1)} .$$

If this last term is strictly greater than 1 then there will be no admissible cycles and, therefore, no equilibria. □

Critical coupling. The condition in Proposition 1.7 is easy to check and provides a way to tune the coupling constants to make the set of admissible cycles empty, and thus phase-locking (from above or from below) impossible. It is now natural to ask what is the minimum value of σ such that there are admissible cycles, to get a sharper necessary condition for the existence of phase-locked configurations.

Definition 1.9 (Critical coupling)**.**

We call *critical coupling* $\sigma_*^{(\pm)}$ for the $(\pm)$ projection the minimum value of σ such that there are admissible cycles ($\mathcal{A}^{(\pm)} \neq \emptyset$).

It follows directly from the definition that $\sigma_s^{(\pm)} < \sigma_*^{(\pm)}$. To find its value, notice first that there can be admissible cycles $x^{(\pm)}$ (Def. 1.6) if and only if

$$\min_{x \in \ker D^{k-1}} \left\| \frac{\beta^{(\pm)}}{\sigma^{\updownarrow}} + x \right\|_\infty \leq 1 \,. \tag{1.48}$$

By manipulating this expression, we can get the exact value of the critical coupling as a solution to a linear optimization problem.

Theorem 1.2 (Value of the critical coupling)**.** *The critical coupling $\sigma_*^{(\pm)}$ can be found in the solution of a linear optimization problem*

$$\sigma_*^{(+)} = \min_{x \in \ker B^{k+1}} \left\| \beta^{(+)} + x \right\|_\infty , \quad \sigma_*^{(-)} = \min_{x \in \ker D^{k-1}} \left\| \beta^{(-)} + x \right\|_\infty , \tag{1.49}$$

which corresponds to the ℓ^∞-norm distance of $\beta^{(\pm)}$ from the space of weighted $(k+1)$-cycles (resp. $(k-1)$-cocycles).

Proof. Using Equation (1.48), we first show that the critical couplings $\sigma_*^{(-)}, \sigma_*^{(+)}$ satisfy, respectively.

$$\min_{x \in \ker D^{k-1}} \left\| \frac{\beta^{(-)}}{\sigma_*^{(-)}} + x \right\|_\infty = 1, \quad \min_{x \in \ker B^{k+1}} \left\| \frac{\beta^{(+)}}{\sigma_*^{(+)}} + x \right\|_\infty = 1 \,. \tag{1.50}$$

If the statement were false and

$$\min_{x \in \ker D^{k-1}} \left\| \frac{\beta^{(-)}}{\sigma_*^{(-)}} + x \right\|_\infty = a \,,$$

with $0 < a < 1$, then we could divide both sides by a and get

$$\min_{x \in \ker D^{k-1}} \left\| \frac{\beta^{(-)}}{a\sigma_*^{(-)}} + \frac{1}{a} x \right\|_\infty = 1 \,,$$

which means that for $\sigma = a\sigma_*^{(-)} < \sigma_*^{(-)}$ there is an admissible cycle $\frac{x}{a}$, which is impossible because we assumed that $\sigma_*^{(-)}$ is the smallest coupling with that property.

Then, multiplying both terms of Equation (1.50) by $\sigma_*^{(-)}$, we have

$$\min_{x \in \ker D^{k-1}} \left\| \beta^{(-)} + \sigma_*^{(-)} x \right\|_\infty = \sigma_*^{(-)} \,.$$

It is now possible to perform a linear change of variable in the optimization problem $\sigma_*^{(-)} x \to \widetilde{x}$ which will change the optimal solution position but not the optimum itself. This means that $\sigma_*^{(-)}$ disappears from the left-hand side and it is found as the solution to the optimization problem above. □

Closed formulae for the critical coupling. In some special cases, the critical coupling admits a closed formula. For example, for the $(-)$ projection of the simplicial Kuramoto dynamics on the edges of a connected simplicial complex, the set of admissible vectors and the critical coupling can both be found explicitly.

Theorem 1.3 (Critical coupling in the edge simplicial Kuramoto)**.** *For the* $(-)$ *component of the edge dynamics on a connected simplicial complex, it holds that*

$$x^{(-)} \in \mathcal{A}^{(-)} \iff x^{(-)} = x\mathbb{1}\,, \tag{1.51}$$

where

$$-\min\left(\frac{\beta^{(-)}}{\sigma^{\downarrow}}\right) - 1 \leq x \leq -\max\left(\frac{\beta^{(-)}}{\sigma^{\downarrow}}\right) + 1\,, \tag{1.52}$$

and

$$\sigma_*^{(-)} = \frac{\max\left(\beta^{(-)}\right) - \min\left(\beta^{(-)}\right)}{2}\,. \tag{1.53}$$

Proof. If the complex is connected we have that D^0 has a 1-dimensional kernel given by span $\{\mathbb{1}\}$. This means that there are admissible vectors if and only if

$$\left\| \frac{\beta^{(-)}}{\sigma^{\downarrow}} + x\mathbb{1} \right\|_{\infty} \leq 1 \iff -1 \leq \frac{\beta_i^{(-)}}{\sigma^{\downarrow}} + x \leq 1\,,$$

$\forall i = 1, \ldots, n_0$, which holds if and only if Equation (1.52) holds, and has solutions only when

$$\max\left(-\frac{\beta^{(-)}}{\sigma^{\downarrow}}\right) - 1 \leq \min\left(-\frac{\beta^{(-)}}{\sigma^{\downarrow}}\right) + 1 \iff \sigma^{\downarrow} \geq \frac{\max(\beta^{(-)}) - \min(\beta^{(-)})}{2}\,.$$

□

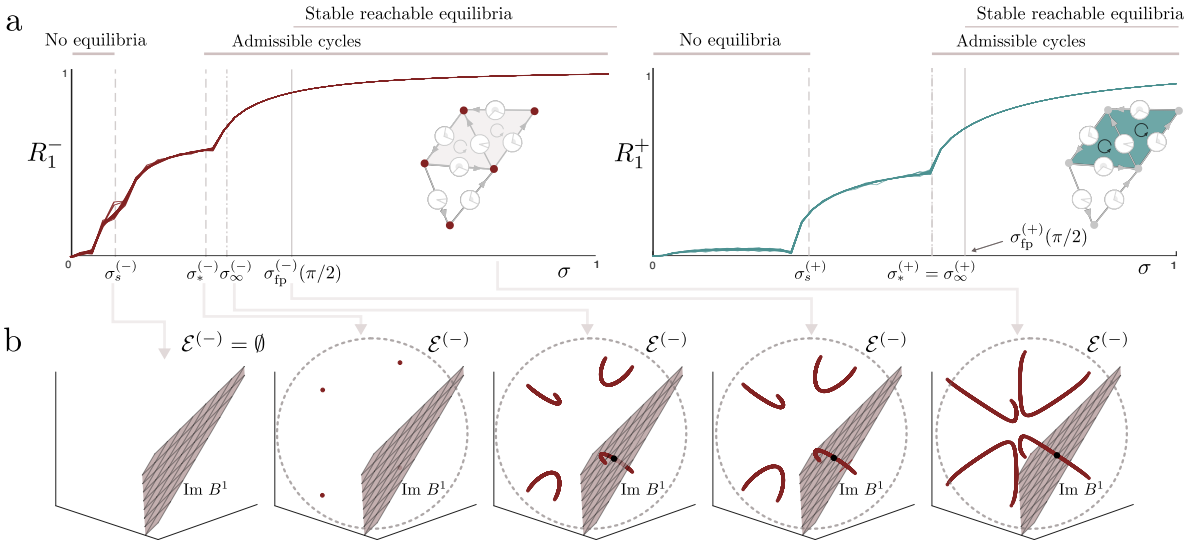

Fig. 1.6 **a.** Fixing the natural frequencies ω, we simulate the edge simplicial Kuramoto model on a small simplicial complex with 20 different initial phase configurations, for values of $\sigma \in [0, 1]$, and compute the time-averaged partial order parameters R_1^- (left), R_1^+ (right) from $t = 0$ to $t = 1000$. The vertical lines correspond to the values of σ_s (Proposition 1.7), σ_* (Theorem 1.2), σ_∞ (Equation (1.61)) and σ_{fp} (Theorem 1.5). If we identify the last "jump" in the order with the emergence of reachable equilibria, then we see how the special values of σ we derived actually bound its value from below and from above. As predicted by Theorem 1.4, the equilibrium transition value for the (+) projection is exactly $\sigma_*^{(+)} = \sigma_\infty^{(+)}$. **b.** The meaning of the different values of σ is depicted by numerically computing the equilibrium set $\mathcal{E}^{(-)} \subset \mathbb{R}^3$ (Definition 1.7 with $m_- = 0$) for the edge dynamics on a 2-simplex. We see how the equilibrium set $\mathcal{E}^{(-)}$ is empty for $\sigma = \sigma_s^{(-)}$, it first appears as a discrete set of points for $\sigma = \sigma_*^{(-)}$ and grows, intersecting the plane $\text{Im}\, B^1$ for $\sigma \geq \sigma_{\text{fp}}^{(-)}$ giving rise to a reachable equilibrium (Definition 1.8), marked here as a black dot.

Asymmetry in the components. It is now worth noting that the properties of the projections $\theta^{(+)} = D^k\theta$ and $\theta^{(-)} = B^k\theta$ are not entirely symmetrical, as the space of $(k+1)$-cycles can be trivial ($\ker B^{k+1} = \{0\}$) and thus there are situations in which the space of admissible cycles is simply $\mathcal{A}^{(+)} = \{0\}$. The same cannot be said for $\ker D^{k-1}$ resulting in the down projection $(-)$ being generally harder to treat.

To shed more light on this, let us consider the k-th order dynamics on a simplicial complex X which has at least one k-simplex. From Definition 1.6, the existence of equilibria for both of them depends on the presence or absence of admissible cycles which, respectively, must belong to $\ker B^{k+1}$ and $\ker D^{k-1}$. The asymmetry stems from the fact that D^{k-1} cannot have a

trivial kernel because

$$\ker D^{k-1} = (\operatorname{Im} B^k)^\perp \underbrace{=}_{\text{Theorem 0.3}} \operatorname{Im} D^{k-2} \oplus \ker L^{k-1}\,, \tag{1.54}$$

and thus

- if $k = 1$, then $\ker D^0 = \ker L^0$, which is non-trivial as there is at least one connected component;
- if $k > 1$ then $\dim \ker D^{k-1} \geq \dim \operatorname{Im} D^{k-2}$ which is nonzero because, by inclusion, there is a nonzero number of $(k-1)$-simplices and D^{k-2} is not an all-zero matrix.

The same cannot be said for B^{k+1} as, in general, there is no restriction on the number of $(k+1)$-cycles. In fact, on the same line of Equation (1.54),

$$\ker B^{k+1} = (\operatorname{Im} D^k)^\perp = \operatorname{Im} B^{k+2} \oplus \ker L^{k+1}\,,$$

which is empty when there are no $(k+2)$-simplices and no $(k+1)$-holes. Therefore, the case of $\ker B^{k+1} = \{0\}$ deserves a special treatment.

Theorem 1.4 (No higher-order cycles)**.** *If there are no $(k+1)$-cycles ($\ker B^{k+1} = \{0\}$) then the following properties hold:*

1. *if $\mathcal{A}^{(+)} \neq \emptyset$ then $\mathcal{A}^{(+)} = \{0\}$;*
2. *if $\mathcal{A}^{(+)} \neq \emptyset$ then the equilibrium set is a discrete set of points given by*

$$\mathcal{E}^{(+)} = \left\{ (-1)^{s_+} \odot \arcsin\left(\frac{\beta^{(+)}}{\sigma^\uparrow}\right) + s_+\pi : s \in \{0,1\}^{n_{k+1}} \right\}\,; \tag{1.55}$$

3. *$\sigma_*^{(+)} = \left\|\beta^{(+)}\right\|_\infty$;*
4. *All equilibria are reachable $\mathcal{E}^{(+)} = \mathcal{R}^{(+)}$.*

Proof. We prove each statement below.
1. It is trivial because 0 is the only vector in $\ker B^{k+1}$.
2. Directly follows from Equation (1.39) with $x^{(+)} = 0$.

3. 0 is the only vector in $\ker B^{k+1}$ so it will be admissible if and only if

$$\left\| \frac{\beta^{(+)}}{\sigma^{\uparrow}} \right\|_{\infty} \leq 1 \,.$$

The smallest value of $\sigma^{\uparrow}$ for which this holds is $\sigma^{\uparrow} = \left\|\beta^{(+)}\right\|_{\infty}$.

According to Definition 1.8, an equilibrium is reachable for the (+) projection if it belongs to $\operatorname{Im} D^k$. In this case,

$$\operatorname{Im} D^k = (\ker(D^k)^*)^{\perp} = (\ker B^{k+1})^{\perp} = \{0\}^{\perp} = \mathbb{R}^{n_{k+1}} \,,$$

which concludes the proof. □

Remark 1.2 (Implications of the absence of $(k+1)$-cycles)**.** *Notice how, in this case, the critical coupling $\sigma_*^{(+)}$ is both the transition value for the existence of admissible cycles (by Definition 1.6) and the existence of the equilibria of the divergence-free component (by Proposition 1.6). This means that $\sigma^{\uparrow} \geq \sigma_*^{(+)}$ is a necessary and sufficient condition for the existence of phase-locked configuration. From this general result, we can obtain* for free *the well-known [76] exact equilibrium transition for the node Kuramoto on trees as, by definition, they have no* 1*-cycles.*

1.4.4 Sufficient conditions for phase locking

Necessary conditions for equilibrium are useful in a setting where we are interested in pushing the system to a *non-equilibrium* state. If, in fact, we are able to tune the coupling strength below one of the bounds derived above (σ_s or σ_*), we are guaranteed that the system will not reach a phase-locked configuration. If, however, we want the system to be phase-locked, we need *sufficient* conditions that can ensure the existence of such equilibria.

An elegant bound on $\sigma^{\updownarrow}$, that both ensures the existence of equilibria and that is easy to compute, can be found generalizing one of the results proven in Ref. [120, Theorem 4.7] by using the proof technique first introduced in Ref. [76] for the node Kuramoto.

Theorem 1.5 (Sufficient condition for the existence of stable reachable equilibria)**.** *For any* $\gamma \in (0, \pi/2)$*, if*

$$\sigma^{\updownarrow} \geq \sigma_{\text{fp}}^{(\pm)}(\gamma) := \frac{\sqrt{\max_i w_i^{k\pm1}}}{\sin(\gamma)} \left\| \beta^{(\pm)} \right\|_{(k\pm1)} , \tag{1.56}$$

there exists an asymptotically stable reachable equilibrium for the $(\pm)$ *dynamics such that*

$$\left\| \theta^{(\pm)} \right\|_{\infty} \leq \gamma . \tag{1.57}$$

Proof. The proof directly follows the constructions in Ref. [76, Theorem 2] by rewriting the equilibrium equation for the projection dynamics as a fixed point equation (hence the subscript *fp* in σ) $x = f(x)$ and finding $\sigma^{\updownarrow}$ such that f is a continuous function from a convex compact set to itself. Brouwer's fixed point theorem then provides the existence of a fixed point (a reachable equilibrium) in this set. The full proof can be found in Section 1.4.5. □

Remark 1.3 (Implications of the sufficient condition)**.** *Four important observations should be highlighted from Theorem 1.5:*

1. *it is always possible to tune the coupling strengths in order for the curl-free and divergence-free components to independently reach equilibrium;*
2. *after a certain value of the coupling strength, these equilibrium configurations always exist and at least one of them is close to the origin;*
3. *increasing the coupling will also decrease the equilibrium's* ℓ^{∞}*-norm;*
4. *we see from the definition of the simplicial order parameter Equation* (1.22) *that, if each component of the projection is close to* 0*, then the configuration will be such that* $R_k \approx 1$ *i.e. phase synchronized (Section 1.3.6).*

We also highlight that, when the complex is unweighted, the expression of the bound becomes

$$\sigma_{\text{fp}}^{(\pm)}(\gamma) = \frac{1}{\sin(\gamma)} \left\| \beta^{(\pm)} \right\|_2 . \tag{1.58}$$

For the node Kuramoto and for $\gamma = \frac{\pi}{2}$, it reduces to

$$\sigma_{\text{fp}}^{(+)} = \left\| B_1^{\dagger} \omega \right\|_2 = \left\| (L^0)^{\dagger} B_1^{\top} \omega \right\|_2 , \tag{1.59}$$

which can be approximated by the well-known bound

$$\sigma \geq \frac{1}{\lambda_2(L^0)} \left\| B_1^\top \omega \right\|_2 , \tag{1.60}$$

where $\lambda_2(L^0)$ is the Fiedler eigenvalue of the network, i.e. the smallest nonzero eigenvalue of L^0.

Another interesting observation is that

$$\left\| \beta^{(+)} \right\|^2_{(k+1)} = \left\langle (B^{k+1})^\dagger \omega, (B^{k+1})^\dagger \omega \right\rangle_{(k+1)} = \left\langle \omega, (B^{k+1} D^k)^\dagger \omega \right\rangle_{(k)} = \left\langle \omega, (L^k_\uparrow)^\dagger \omega \right\rangle_{(k)} ,$$

which is exactly the **effective resistance** of ω as defined in Ref. [121].

Remark 1.4 (Intepretation of the sufficient condition)**.** *To have equilibrium, the coupling must overcome the "structural" resistance of the simplicial complex, encoded in both the incidence structure ($L^k_\updownarrow$) and the natural frequencies. This is a powerful observation because it means that it might be possible to define pairs of structures and frequencies to reach particular types of dynamics or control the frequencies to move across regimes.*

Finally, while Theorem 1.5 ensures the existence of reachable equilibria, in practice its value $\sigma^{(\pm)}_{\text{fp}}$ tends to be conservative and to overestimate the minimum value of σ for which stable reachable equilibria exist. In perfect analogy with the node Kuramoto literature [119], it is often seen in practice that the infinity-norm of the natural potential

$$\sigma^{(\pm)}_\infty := \left\| \beta^{(\pm)} \right\|_\infty \tag{1.61}$$

is closer to the true reachability threshold and thus provides a sharper bound. This value, moreover, exactly coincides with the reachability transition in some special cases, such as in the one of Theorem 1.4.

The different bounds on σ found in Sections 1.4.3 and 1.4.4 are shown in Figure 1.6, where they are related to the partial order parameters on a small simplicial complex. We see how σ_* and σ_{fp} actually bound the point of the last jump, corresponding to the transition value after which the dynamics admits reachable equilibria.

1.4.5 Proof of Theorem 1.5

We show here the full proof of Theorem 1.5, which is quite technical due to the necessity to keep track of the weights and all the subspaces involved. For simplicity, we prove the result only for the $(-)$ component. The $(+)$ case can be easily recovered by replacing $L_{\uparrow}^{k-1}$ with $L_{\downarrow}^{k+1}$, $\omega^{(-)}$ with $\omega^{(+)}$ and $\sigma^{\downarrow}$ with $\sigma^{\uparrow}$.

The idea of the proof, inspired by Ref. [76], is in itself simple. We restrict the dynamics to the reachable subspace $\operatorname{Im} L_{\uparrow}^{k-1}$ by finding the evolution equations of its coefficients w.r.t. a basis, rewrite its equilibrium equation as a fixed-point equation, and then apply Brouwer's fixed-point theorem to ensure the presence of a solution.

Restricting the dynamics to the reachable subspace. First, as $L_{\uparrow}^{k-1}$ is the matrix representation of a self-adjoint, positive semidefinite operator, we can consider its eigendecomposition (Proposition 0.5)

$$L_{\uparrow}^{k-1} = V \Lambda V^{\top} W_{k-1}^{-1}\,, \tag{1.62}$$

where Λ is diagonal with non-negative elements and the columns of V are a basis of $\mathbb{R}^{n_{k-1}}$ that is orthonormal w.r.t. W_{k-1}^{-1}, i.e. $V^{\top} W_{k-1}^{-1} V = I_{n_{k-1}}$. Let us assume that the zero eigenvalues of Λ are the last ones in the diagonal, so that

$$\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_r, 0, \dots, 0)\,,$$

where $r = \operatorname{rank}(L_{\uparrow}^{k-1})$.

We want to restrict the dynamics to the subspace $\operatorname{Im} L_{\uparrow}^{k-1}$. To do that, we drop the columns associated with the zero eigenvalues (which span $\ker L_{\uparrow}^{k-1}$) and consider the compact eigendecomposition

$$L_{\uparrow}^{k-1} = \widetilde{V} \widetilde{\Lambda} \widetilde{V}^{\top} W_{k-1}^{-1}\,, \tag{1.63}$$

$\widetilde{V}$ consists of the first r columns of V and $\widetilde{\Lambda} = \operatorname{diag}(\lambda_1, \dots, \lambda_r)$. The columns of $\widetilde{V}$ form an orthonormal basis of the reachable subspace $\operatorname{Im} L_{\uparrow}^{k-1}$ and thus $\widetilde{V}^{\top} W_{k-1}^{-1} \widetilde{V} = I_r$

We can now rewrite the simplicial Kuramoto dynamics of the $(-)$ projection (Equation (1.31)) in the basis $\widetilde{V}$, $\theta^{(-)} = \widetilde{V}c$ for some vector of coefficients $c \in \mathbb{R}^r$:

$$\frac{d}{dt}\widetilde{V}c = \omega^{(-)} - \sigma^{\downarrow} L_{\uparrow}^{k-1} \sin(\widetilde{V}c) = \omega^{(-)} - \sigma^{\downarrow}\widetilde{V}\widetilde{\Lambda}\widetilde{V}^{\top} W_{k-1}^{-1} \sin(\widetilde{V}c)\,. \tag{1.64}$$

With this formulation, we are naturally restricting $\theta^{(-)}$ to lie in the reachable subspace. We find the dynamics of the coefficients c by left multiplying by $\widetilde{V}^{\top} W_{k-1}^{-1}$ and using the fact that $\widetilde{V}^{\top} W_{k-1}^{-1}\widetilde{V} = I_r$

$$\dot{c} = \widetilde{V}^{\top} W_{k-1}^{-1}\omega^{(-)} - \sigma^{\downarrow}\widetilde{\Lambda}\widetilde{V}^{\top} W_{k-1}^{-1} \sin(\widetilde{V}c), \tag{1.65}$$

The coefficients c are associated to a reachable equilibrium configuration if and only if $\dot{c} = 0$, i.e.

$$\widetilde{V}^{\top} W_{k-1}^{-1}\frac{\omega^{(-)}}{\sigma^{\downarrow}} = \widetilde{\Lambda}\widetilde{V}^{\top} W_{k-1}^{-1} \sin(\widetilde{V}c)\,. \tag{1.66}$$

Equilibrium as a fixed point equation. We want to reduce this equation to a fixed point equation, of the form $f(c) = c$ for some function f. First, we write

$$\widetilde{V}^{\top} W_{k-1}^{-1}\frac{\omega^{(-)}}{\sigma^{\downarrow}} = \widetilde{\Lambda}\widetilde{V}^{\top} W_{k-1}^{-1} \sin(\widetilde{V}c) \iff \widetilde{\Lambda}^{-1}\widetilde{V}^{\top} W_{k-1}^{-1}\frac{\omega^{(-)}}{\sigma^{\downarrow}} = \widetilde{V}^{\top} S(c)\widetilde{V}c$$

where we defined $S(c) := W_{k-1}^{-1}\operatorname{diag}(\operatorname{sinc}(\widetilde{V}c))$ with $\operatorname{sinc}(x) = \sin(x)/x$. We then have the fixed point equation

$$c = (\widetilde{V}^{\top} S(c)\widetilde{V})^{-1}\widetilde{\Lambda}^{-1}\widetilde{V}^{\top} W_{k-1}^{-1}\frac{\omega^{(-)}}{\sigma^{\downarrow}} := f(c)\,, \tag{1.67}$$

which makes sense only if the matrix $\widetilde{V}^{\top} S(c)\widetilde{V}$ is invertible.

Lemma 1.2 (Invertibility of $\widetilde{V}^{\top} S(c)\widetilde{V}$)**.** *If $S(c)$ has strictly positive elements, then $\widetilde{V}^{\top} S(c)\widetilde{V}$ is invertible.*

Proof. First, notice that $S(c)$ is a diagonal matrix with positive diagonal elements. We can thus take its square root and get the following

$$\widetilde{V}^{\top} S(c)\widetilde{V} = (S^{\frac{1}{2}}(c)\widetilde{V})^{\top}(S^{\frac{1}{2}}(c)\widetilde{V}) := A^{\top}A.$$

Given that $\ker A^\top = (\operatorname{Im} A)^\perp$, we deduce that $A^\top A$ is invertible if and only if $A = S^{\frac{1}{2}}(c)\widetilde{V}$ has trivial kernel. Moreover, we know that the columns of $\widetilde{V}$ are a basis and thus $\widetilde{V}c = 0 \iff c = 0$. If $S^{\frac{1}{2}}(c)$ is invertible, which holds when $S(c)$ has positive diagonal elements, then its kernel will be trivial and, by extension, the same will hold for A. □

This result on the invertibility of $\widetilde{V}^\top S(c)\widetilde{V}$ hence translates to a condition on $S(c)$.

Lemma 1.3 (Positive definiteness of S)**.** *For any $\gamma \in (0, \pi/2)$, if the coefficients c are such that*

$$\left\|\widetilde{V}c\right\|_{(k-1)} \leq \frac{\gamma}{\sqrt{\max_i w_i^{k-1}}}, \tag{1.68}$$

then $S(c)$ has positive diagonal elements.

Proof. Under the hypothesis of the lemma it holds that

$$\left\|\widetilde{V}c\right\|_\infty \leq \left\|\widetilde{V}c\right\|_2 = \sqrt{\sum_i (\widetilde{V}c)_i^2} = \sqrt{\sum_i w_i^{k-1} \frac{1}{w_i^{k-1}} (\widetilde{V}c)_i^2} \leq \sqrt{\max_i w_i^{k-1}} \left\|\widetilde{V}c\right\|_{(k-1)} \leq \gamma, \tag{1.69}$$

meaning that every component of $\widetilde{V}c$ will belong to the interval $[-\gamma, \gamma]$. The sinc function, which is applied component-wise to $\widetilde{V}c$, is strictly positive in $[-\gamma, \gamma]$ when $\gamma \in (0, \pi/2)$, hence the positive definiteness of $S(c) = W_{k-1}^{-1} \operatorname{diag}(\operatorname{sinc}(\widetilde{V}c)))$. □

We now want to prove that the left-hand side of the equilibrium fixed point eq. (1.67) is a continuous map from the set

$$\mathcal{B} = \left\{ c : \|c\|_2 = \left\|\widetilde{V}c\right\|_{(k-1)} \leq \frac{\gamma}{\sqrt{\max_i w_i^{k-1}}} \right\} \tag{1.70}$$

to itself. First, one has the following inequality.

$$\|f(c)\|_2 = \left\| (\widetilde{V}^\top S(c) \widetilde{V})^{-1} \widetilde{\Lambda}^{-1} \widetilde{V}^\top W_{k-1}^{-1} \frac{\omega^{(-)}}{\sigma^\downarrow} \right\|_2 \tag{1.71}$$

$$\leq \frac{1}{\sigma^\downarrow} \left\| (\widetilde{V}^\top S(c) \widetilde{V})^{-1} \right\|_2 \left\| \widetilde{\Lambda}^{-1} \widetilde{V}^\top W_{k-1}^{-1} \omega^{(-)} \right\|_2, \tag{1.72}$$

where the first term is the matrix norm. Let us look at the two terms of eq. (1.71) separately, starting from the right one.

Lemma 1.4. *If $c \in \mathcal{B}$ then*

$$\left\|\widetilde{\Lambda}^{-1}\widetilde{V}^\top W_{k-1}^{-1}\omega^{(-)}\right\|_2 = \left\|\beta^{(-)}\right\|_{(k-1)}. \tag{1.73}$$

Proof.

$$\begin{aligned}\left\|\widetilde{\Lambda}^{-1}\widetilde{V}^\top W_{k-1}^{-1}\omega^{(-)}\right\|_2 &= \left\|\underbrace{\widetilde{V}^\top W_{k-1}^{-1}\widetilde{V}}_{I_r}\widetilde{\Lambda}^{-1}\widetilde{V}^\top W_{k-1}^{-1}\omega^{(-)}\right\|_2 \\ &= \left\|\widetilde{V}^\top W_{k-1}^{-1}(L_\uparrow^{k-1})^\dagger\omega^{(-)}\right\|_2 = \left\|\widetilde{V}^\top W_{k-1}^{-1}\beta^{(-)}\right\|_2.\end{aligned}$$

Moreover,

$$\left\|\widetilde{V}^\top W_{k-1}^{-1}\beta^{(-)}\right\|_2^2 = \left\langle\widetilde{V}^\top W_{k-1}^{-1}\beta^{(-)}, \widetilde{V}^\top W_{k-1}^{-1}\beta^{(-)}\right\rangle_2 = \left\langle\widetilde{V}\widetilde{V}^\top W_{k-1}^{-1}\beta^{(-)}, \beta^{(-)}\right\rangle_{(k-1)},$$

but, as $\widetilde{V}\widetilde{V}^\top W_{k-1}^{-1}$ is the orthogonal projection operator onto $\operatorname{Im} L_\uparrow^{k-1}$ and $\beta^{(-)} \in \operatorname{Im} L_\uparrow^{k-1}$, $\widetilde{V}\widetilde{V}^\top W_{k-1}^{-1}\beta^{(-)} = \beta^{(-)}$, and we get our result. □

Let us now analyze the first term of Equation (1.71) and bound it from above.

Lemma 1.5. *If $c \in \mathcal{B}$ then*

$$\left\|(\widetilde{V}^\top S(c)\widetilde{V})^{-1}\right\|_2 \leq \frac{\gamma}{\sin(\gamma)} \tag{1.74}$$

Proof. When $c \in \mathcal{B}$ then Lemma 1.3 tells us that $S(c)$ is positive definite and therefore we can write

$$\left\|(\widetilde{V}^\top S(c)\widetilde{V})^{-1}\right\|_2 = \left\|(A^\top A)^{-1}\right\|_2,$$

with $A = S^{\frac{1}{2}}(c)\widetilde{V}$, for which it holds that

$$\frac{1}{\left\|(A^\top A)^{-1}\right\|_2} = \min_{\|v\|_2=1}\left\|A^\top Av\right\|_2. \tag{1.75}$$

We apply here the Cauchy-Schwarz inequality

$$\left|\left\langle A^\top Ac, v\right\rangle\right| \leq \left\|A^\top Av\right\|\|v\|$$

and find that the right-hand side of eq. (1.75) can be bounded from below

$$
\begin{aligned}
\min_{\|v\|_2=1} \left\|A^\top A v\right\|_2 &\geq \min_{\|v\|_2=1} \left|\left\langle A^\top A v, v\right\rangle_2\right| = \min_{\|v\|_2=1} \|Av\|_2^2 && \text{(definition of } A\text{)} \\
&= \left(\min_{\|v\|_2=1} \left\|S^{\frac{1}{2}}(c)\widetilde{V}v\right\|_2\right)^2 && \text{(definition of } S(c)\text{)} \\
&= \left(\min_{\|\widetilde{V}v\|_{(k-1)}=1} \left\|\operatorname{diag}(\operatorname{sinc}(\widetilde{V}c))^{\frac{1}{2}}\widetilde{V}v\right\|_{(k-1)}\right)^2 && \text{(The columns of } \widetilde{V} \text{ span Im } L_\uparrow^{k-1}\text{)} \\
&= \left(\min_{\|\theta\|_{(k-1)}=1, \theta\in \operatorname{Im} L_\uparrow^{k-1}} \left\|\operatorname{diag}(\operatorname{sinc}(\widetilde{V}c))^{\frac{1}{2}}\theta\right\|_{(k-1)}\right)^2 && \text{(Remove a constraint)} \\
&\geq \left(\min_{\|\theta\|_{(k-1)}=1} \left\|\operatorname{diag}(\operatorname{sinc}(\widetilde{V}c))^{\frac{1}{2}}\theta\right\|_{(k-1)}\right)^2 && \text{(Equation (1.75))} \\
&= \left(\left\|\operatorname{diag}(\operatorname{sinc}(\widetilde{V}c))^{-\frac{1}{2}}\right\|_{(k-1)}^2\right)^{-1}.
\end{aligned}
$$

We have proven that

$$
\frac{1}{\left\|(A^\top A)^{-1}\right\|_2} \geq \frac{1}{\left\|\operatorname{diag}(\operatorname{sinc}(\widetilde{V}c))^{-\frac{1}{2}}\right\|_{(k-1)}^2},
$$

or, equivalently,

$$
\left\|(\widetilde{V}^\top S(c)\widetilde{V})^{-1}\right\|_2 = \left\|(A^* A)^{-1}\right\|_2 \leq \left\|\operatorname{diag}(\operatorname{sinc}(\widetilde{V}c))^{-\frac{1}{2}}\right\|_{(k-1)}^2. \tag{1.76}
$$

This term can be further rewritten by trying to explicitly compute the weighted $(k-1)$-norm of a matrix. Let D be a diagonal matrix. By the definition of operator norm

$$
\begin{aligned}
\|D\|_{(k-1)} &= \max_{\|x\|_{(k-1)}=1} \langle Dx, x\rangle_{(k-1)} = \max_{x^\top W_{k-1}^{-1} x=1} x^\top D W_{k-1}^{-1} x \quad \text{Change variable } y := W_{k-1}^{-1/2} x \\
&= \max_{\|y\|_2=1} y^\top W_{k-1}^{\frac{1}{2}} D W_{k-1}^{-\frac{1}{2}} y = \max_{\|y\|_2=1} y^\top D y \\
&= \|D\|_2 = \max_i |D_{ii}|.
\end{aligned}
$$

Therefore, for diagonal matrix and diagonal inner product matrices, the $(k-1)$-norm is equal to the standard matrix 2-norm. In our case, this means that

$$\left\|\mathrm{diag}(\mathrm{sinc}(\widetilde{V}c))^{-\frac{1}{2}}\right\|^2_{(k-1)} = \left\|\mathrm{diag}(\mathrm{sinc}(\widetilde{V}c)^{-\frac{1}{2}}\right\|^2_2 = \max_i \left|\frac{1}{\mathrm{sinc}(\widetilde{V}c)_i}\right|,$$

because $S(c)$ is diagonal with positive diagonal elements. We now remove the dependency on c by taking a maximum over $\mathcal{B}$

$$\left\|\mathrm{diag}(\mathrm{sinc}(\widetilde{V}c))^{-\frac{1}{2}}\right\|^2_{(k-1)} \leq \max_{c\in\mathcal{B}} \max_i \left|\frac{1}{\mathrm{sinc}(\widetilde{V}c)_i}\right| \leq \max_{x\in[-\gamma,\gamma]} \frac{1}{\mathrm{sinc}(x)} = \frac{\gamma}{\sin(\gamma)}. \tag{1.77}$$

Putting everything together, we have that

$$\left\|(\widetilde{V}^\top S(c)\widetilde{V})^{-1}\right\|_2 \leq \left\|\mathrm{diag}(\mathrm{sinc}(\widetilde{V}c)))^{-\frac{1}{2}}\right\|^2_{(k-1)} \leq \frac{\gamma}{\sin(\gamma)}\,.$$

□

Applying lemmas 1.4 and 1.5 to eq. (1.71), we have

$$\|f(c)\|_2 = \left\|(\widetilde{V}^\top S(c)\widetilde{V})^{-1}\widetilde{\Lambda}^{-1}\widetilde{V}^\top W_{k-1}^{-1}\frac{\omega^{(-)}}{\sigma^\downarrow}\right\|_2 \leq \frac{1}{\sigma^\downarrow}\frac{\gamma}{\sin(\gamma)}\left\|\beta^{(-)}\right\|_{(k-1)}\,. \tag{1.78}$$

We now ask for a sufficient condition for f to map the ball $\mathcal{B}$ to itself. $f(c)\in\mathcal{B}$ if and only if

$$\|f(c)\|_2 \leq \frac{\gamma}{\sqrt{\max_i(w_i^{k-1})}} \tag{1.79}$$

which holds if

$$\frac{1}{\sigma^\downarrow}\frac{\gamma}{\sin(\gamma)}\left\|\beta^{(-)}\right\|_{(k-1)} \leq \frac{\gamma}{\sqrt{\max_i(w_i^{k-1})}} \iff \sigma^\downarrow \geq \frac{\sqrt{\max_i(w_i^{k-1})}}{\sin(\gamma)}\left\|\beta^{(-)}\right\|_{(k-1)}\,. \tag{1.80}$$

This proves that, under the condition of the theorem, $f(c)$ maps the closed ball $\mathcal{B}$ to itself, and so Brouwer's theorem ensures the existence of a fixed point (i.e. a reachable equilibrium) $\theta^{(-)}_{\mathrm{eq}} = \widetilde{V}c_{\mathrm{eq}}$ with $c_{\mathrm{eq}}\in\mathcal{B}$.

Asymptotic stability. The asymptotic stability of $\theta_{\mathrm{eq}}^{(-)}$ can be shown by computing the Jacobian of the reachable dynamics eq. (1.65)

$$J^{(-)}(c_{\mathrm{eq}}) = -\sigma^{\downarrow}\widetilde{\Lambda}\widetilde{V}^{\top}W_{k-1}^{-1}\,\mathrm{diag}(\cos(\theta_{\mathrm{eq}}^{(-)}))\widetilde{V}\,, \tag{1.81}$$

which has the same nonzero eigenvalues as the similar matrix

$$\widetilde{J}^{(-)}(c_{\mathrm{eq}}) = -\sigma^{\downarrow}\widetilde{\Lambda}^{\frac{1}{2}}\widetilde{V}^{\top}W_{k-1}^{-1}\,\mathrm{diag}(\cos(\theta_{\mathrm{eq}}^{(-)}))\widetilde{V}\widetilde{\Lambda}^{\frac{1}{2}}.$$

If $c_{\mathrm{eq}} \in \mathcal{B}$ and $\gamma \in (0, \pi/2)$, then

$$\left\|\theta_{\mathrm{eq}}^{(-)}\right\|_{(k-1)} \leq \frac{\gamma}{\sqrt{\max_i(w_i^{k-1})}} \implies \left\|\theta_{\mathrm{eq}}^{(-)}\right\|_{\infty} \leq \gamma \implies \cos\left(\theta_{\mathrm{eq}}^{(-)}\right) > 0\,,$$

and thus

$$\widetilde{J}^{(-)}(c_{\mathrm{eq}}) = -\sigma^{\downarrow}\left(\mathrm{diag}\left(\cos\left(\theta_{\mathrm{eq}}^{(-)}\right)\right)^{\frac{1}{2}}W_{k-1}^{-\frac{1}{2}}\widetilde{V}\widetilde{\Lambda}^{\frac{1}{2}}\right)^{\top}\left(\mathrm{diag}\left(\cos\left(\theta_{\mathrm{eq}}^{(-)}\right)\right)^{\frac{1}{2}}W_{k-1}^{-\frac{1}{2}}\widetilde{V}\widetilde{\Lambda}^{\frac{1}{2}}\right) = -\sigma^{\downarrow}A^{\top}A\,,$$

which is negative definite as $A := \mathrm{diag}\left(\cos(\theta_{\mathrm{eq}}^{(-)})\right)^{\frac{1}{2}}W_{k-1}^{-\frac{1}{2}}\widetilde{V}\widetilde{\Lambda}^{\frac{1}{2}}$ has a trivial kernel and $\sigma^{\downarrow} > 0$.

1.5 Application to functional connectivity reconstruction

Neural oscillations and frequencies are fundamental to our understanding of brain function [122–125], and contributes to the emergence of brain functional networks [126]. In this context, oscillator models, and the node Kuramoto model in particular, have been extensively used in computational neuroscience, as they offer simple yet powerful and flexible phenomenological framework for studying simplified versions of the dynamics of neuronal or brain networks with various degrees of complexity [127, 69, 128–131]. By treating neurons or brain regions as oscillators that interact with each other, these models can capture significant features of brain activity observed in experiments, such as the presence of rhythms and oscillations. While oscillator models have been widely used to study brain dynamics, it is important to note that most of these have focused on pairwise interactions between neurons or brain regions. This is due in part to the fact that pairwise interactions are simpler to model or analyze and that anatomically it is more realistic to consider the dynamics taking place on networks rather than higher-order systems. However, recent studies have suggested that higher-order

interactions may also play an important role in brain dynamics, both functionally [18, 132] and structurally [92, 133]. These interactions involve three or more elements and can give rise to emergent phenomena that cannot be explained by pairwise interactions alone. Given the potential importance of these higher-order interactions, it is natural to apply models of higher-order synchronization to brain data. These models might offer a more comprehensive framework for studying the dynamics of large-scale brain networks and have the potential to give us new insights into the mechanisms underlying cognition and behavior.

To test this hypothesis, we study how well simplicial Kuramoto models can reproduce brain correlation patterns. Following the methodology proposed in Refs. [129, 130], we run simulations of three different variants of the simplicial Kuramoto model on a real structural connectome, the network that describes the connectivity structures between regions of the human cerebral cortex, and we investigate how well each model can reproduce the resting-state functional activity experimentally measured.

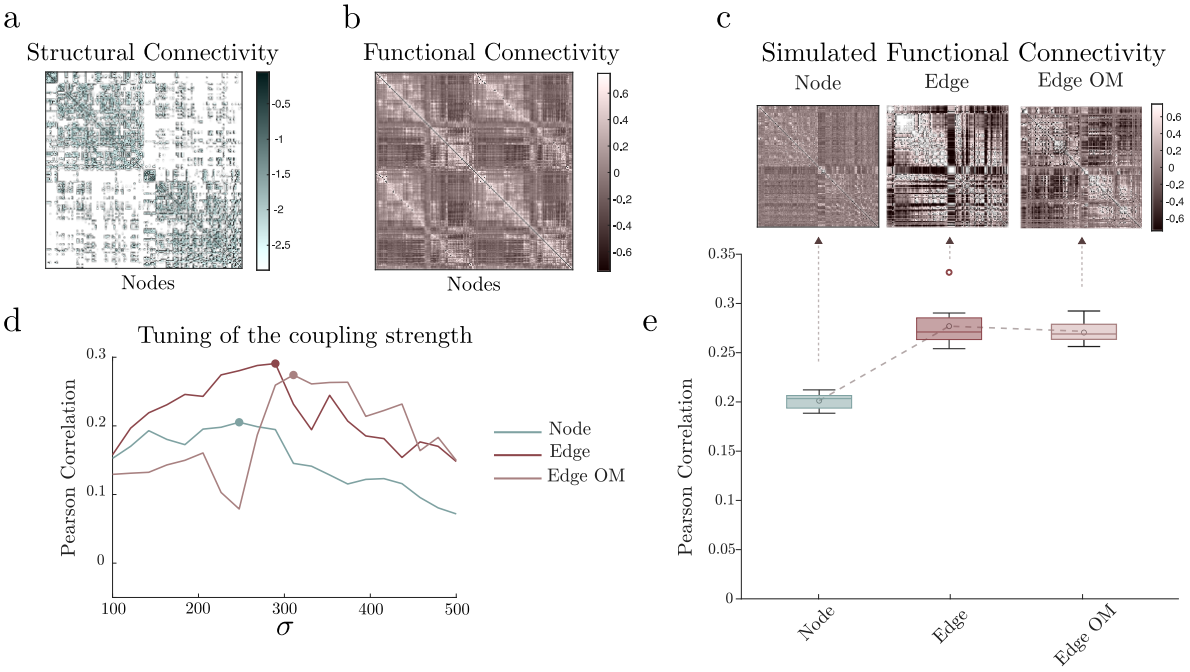

Fig. 1.7 **a.** Structural connectivity matrix representing the weighted network onto which we simulate the dynamics. The color represents the logarithm of the weight. **b.** The empirical functional connectivity matrix. **c.** We simulate three different variants of the simplicial Kuramoto model and compute the correlation matrices of their post-processed trajectories as simulated FC matrices. **d.** The coupling strength σ is tuned for each model by scanning 20 values between 100 and 500. **e.** Pearson correlations between the empirical FC and the simulated FCs for the three models, over 10 simulations.

The structural connectome is encoded in a group-averaged weighted structural connectivity matrix (Figure 1.7a), obtained by diffusion imaging and tractography, by parcellating the brain into $N = 200$ regions, which here take the role of nodes connected by $M = 6040$ weighted edges. From the network adjacency matrix, we derive the incidence matrix $B_1 \in \{-1, 1\}^{N\times M}$ by choosing randomly edges orientations. To achieve consistency with Ref. [129], the connection weights $K_1, \dots, K_M$ are included, after being inverted, as weights on the edges $W_1 = \mathrm{diag}(\frac{1}{K_1}, \dots, \frac{1}{K_M})$ and the tract lengths are encoded in an edge frustration vector $\alpha \in \mathbb{R}^M$. The natural frequencies for both node-based and edge-based models are sampled independently from a Gaussian distribution with a mean of $2\pi\, 40$ and a standard deviation of $2\pi\, 0.1$. We compare the following three models:

1. *Node Kuramoto-Sakaguchi model (Node).* This is, by construction, the classical Kuramoto-Sakaguchi model of Equation (1.3) with added edge frustrations α_{ij}, used in Ref. [129]

$$\dot{\theta}_i = \omega_i - \sigma \sum_{j=1}^{200} K_{ij} \sin(\theta_i - \theta_j + \alpha_{ij}) \,. \tag{1.82}$$

2. *Edge Simplicial Kuramoto (Edge).* The simplest possible simplicial Kuramoto model defined on the edges

$$\dot{\theta} = \omega - \sigma B_1^\top \sin(B_1 W_1^{-1} \theta) \,. \tag{1.83}$$

In addition to the simplicial Kuramoto model we described in this chapter, Ref. [33] proposes to modulate the interaction terms by multiplying them by the order parameters of the model. They show that this model has an interesting, rich behavior, displaying an explosive synchronization transition in the coupling strength. That model, however, cannot be directly applied here as it requires the presence of triangles, which we do not include in our picture. As a proxy for its behavior, we propose the similar *order-modulated model* (OM), derived by multiplying σ by the order parameter. In other words, the OM model is the gradient system of the square order parameter.

$$\dot{\theta} = \omega + \frac{1}{2} C_k W_k \nabla_\theta R_k^2(\theta) \,, \tag{1.84}$$

that results in:

3. the *Order-modulated edge simplicial Kuramoto (Edge OM)*

$$\dot{\theta} = \omega - \sigma R_1(\theta) B_1^\top \sin(B_1 W_1^{-1} \theta)\,. \tag{1.85}$$

Models 2, 3 are defined on the edges of the network, but, given that we want to reproduce a node-wise functional connectivity matrix, we consider the projections of their phases onto the nodes $\theta^{(-)}$ to get node-wise trajectories. For this reason, note that it is not necessary to numerically solve all the M equations on the edges, but it is enough to directly integrate the projected dynamics.

1.5.1 Simulations

Following Ref. [129], the simulations are run for a total of $T = 812$ seconds with a time resolution of $\Delta t = 1$ms (using MATLAB ode45), and the first 20 seconds are discarded to allow the dynamics to reach stationarity. We then take the trajectories, convert them into downsampled BOLD signals, filter them with a lowpass cutoff of $c = 0.25$Hz, and use them to compute $N \times N$ pairwise Pearson correlation matrices. These simulated functional connectivity matrices (Figure 1.7b) are then compared to the experimental resting-state functional connectivity (FC) matrix (Figure 1.7c) using Pearson correlation (by correlating the vectorized upper triangular matrix). We repeat this process multiple times for each model by varying the coupling strength in order to tune it. We scan 20 σ values ranging from 100 to 500, and select the optimal one w.r.t Pearson correlation (Figure 1.7e). Given the optimal coupling strength for each model, we then perform 10 simulations for each one of them with different random starting phases and natural frequencies and confront them with the empirical FC matrix. The results are shown in Figure 1.7d, where it is easy to see how the two non-frustrated edge-based models outperform the node ones, achieving an average of $r = 0.27$ correlation against the $r = 0.2$ of the standard node Sakaguchi-Kuramoto. The result is statistically confirmed by an ANOVA test, which achieves p-values lower than 10^{-3}. The effect size against the node Kuramoto model is 0.0757 for the edge model and 0.0692 for the edge OM.

Our findings suggest that an edge-based description of the dynamics might provide a better fit to the experimental data, both outperforming the node-based models and without resorting to additional parameters or internal mechanisms, as for example edge flickering [130] (which was shown to obtain a slightly lower correlation than our edge Kuramoto model). In fact, arguably edge-based simplicial Kuramoto models might provide a better fit to the observed FC correlation structure exactly because the variables are defined on the connections

that link different nodes together, rather than on the nodes themselves [134]. That is, the observed activity of brain regions might be better explained as the result of the information integration taking place via the structural fibers linking the regions, rather than by looking at the brain regions in themselves [135, 136], and display interesting parallels with neural frequency mixing behaviors [137, 138]. Naturally, these results are preliminary and intended as a simple demonstration of the potential of simplicial (and more generally, higher-order) oscillator models in the context of computational neurobiological models.

We refer the interested reader to Ref. [54] for a more comprehensive evaluation of simplicial Kuramoto models on the same tasks.

1.6 Summary and Outlook

In this chapter, we have focused on the analytical study of the standard simplicial Kuramoto model, in which phase oscillators are defined on simplices of a fixed order and interact through the simplicial structure of an underlying complex. By exploiting tools from topology and discrete differential geometry, such as boundary operators and the Hodge Laplacian, we provided a clear mathematical framework to characterize the model as a gradient system and to analyze its synchronization properties.

The simplicity of the model allowed us to derive general results on the structure of its equilibria and on the role played by the coupling strength in determining their stability and reachability. In particular, we were able to obtain necessary and sufficient conditions on the coupling strength for synchronization and to describe how the space of equilibria changes as a function of the coupling parameter. Moreover, we showed that, when the underlying simplicial complex is manifold-like, the simplicial Kuramoto model is strictly equivalent to a standard Kuramoto model defined on an effective pairwise network. This equivalence highlights that genuinely new dynamical behaviors induced by simplicial interactions emerge only on non-manifold-like complexes, where the higher-order structure cannot be reduced to pairwise couplings.

Beyond this equivalence result, we established an explicit connection between simplicial Kuramoto models and another important class of higher-order synchronization models, namely node-based Kuramoto models with interactions defined on hypergraphs. We showed that a simplicial Kuramoto model on a generic complex can be rewritten as a node Kuramoto model with group interactions occurring on an effective dual hypergraph. This mapping reveals both

similarities and crucial differences between the two approaches. In particular, the coupling functions induced by simplicial dynamics do not generally vanish when all phases are equal, implying that the fully synchronized one-cluster state is not guaranteed to exist and that the equations are not invariant under uniform phase shifts. Instead, the dynamics exhibit an invariance associated with the harmonic space of the simplicial complex, which can be larger than the symmetry generated by uniform phase shifts alone.

We further interpreted this peculiarity from the perspective of resonance. Specifically, we showed that the simplicial Kuramoto coupling functions can be derived from a general system of higher-order interacting oscillators operating in a near-resonant regime, under the assumption that the natural frequencies lie in the harmonic space of the complex. This resonance condition directly links the dynamical properties of the model to the topology of the underlying simplicial structure. Although this derivation was carried out explicitly only for the standard simplicial Kuramoto model, it suggests a broader interpretation of simplicial interactions as effective descriptions of resonant higher-order dynamics.

We emphasize that the standard simplicial Kuramoto model analyzed in this chapter represents only one instance within a broader family of simplicial synchronization models. We note that several variants can be obtained by modifying the interaction structure to induce adaptive couplings [33], frustrations [50] or extending the dynamics to cell complexes [103, 139]. Of particular interest, the topological synchronization model based on the Dirac operator on simplicial complexes allows the definition of multi-order signals that are coupled within and across orders [133, 97, 96, 140, 141]. While these extensions are not treated in this chapter, a number of them are discussed and systematically analyzed in the main paper from which this chapter is derived [54].

Finally, while the results presented here focus on synchronization, the formalism naturally extends to other dynamical processes, such as consensus or diffusion, and to more general topological structures beyond simplicial complexes. Overall, this work provides a rigorous analytical foundation for simplicial Kuramoto models and clarifies their relationship with both classical network-based synchronization and hypergraph-based higher-order interactions, offering a starting point for further theoretical developments and applications.

Back to Table of Contents

Chapter 2

Higher-order Laplacian Renormalization

Contents

2.1 Introduction

The renormalization group (RG) [55] is a cornerstone of modern theoretical physics, as it allows us to study how a physical system depends on the scale of observation, defining universality classes and, importantly, formalizing the concept of scale-invariance. It is based on the idea that understanding the behavior of a physical system at a large enough scale doesn't require the knowledge of all the microscopic interacting components it is made of. A coarse, aggregate view is often enough.

While the RG has been a powerful tool for understanding a broad class of physical systems, extending its framework to complex networks has posed a recent and significant challenge.

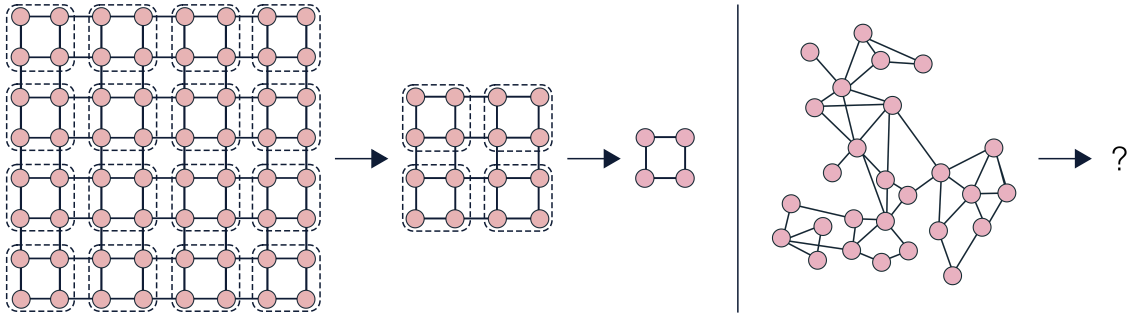

Fig. 2.1 Kadanoff-like block nodes can be easily defined on a regular lattice but are not trivially defined on a heterogeneous network.

In its classical Kadanoff formulation [142], the renormalization group is defined on regular homogeneous lattices, where it is easy to group nodes into blocks that can be coarse-grained into macro-nodes (Figure 2.1).

However, real-world networks rarely exhibit interaction patterns that possess the regularity of lattices. They are heterogeneous, sparse objects where nodes can have different degrees, and long-range connections make all nodes just a few links from one another. There, it is less clear how to appropriately find the notion of locality required for creating Kadanoff-like blocks (Figure 2.1). Nonetheless, extending the RG to networks has gained substantial attention in recent years [56–63, 143] due to its potential to provide insights into the multiscale structural organization of complex networks.

First attempts take the geodesic (shortest-path) length as the natural metric between nodes and coarse grains sets of nodes whose distance is lower than a given scale parameter through a "box-covering" procedure [56]. This has resulted in notable results in unveiling the fractal-like self-similar structure of certain classes of networks [57, 144, 145]. However, most real-world networks possess the *small-world* property [12], meaning that their diameter is extremely small. This turns out to be a severe hindrance to box-covering renormalization methods, where blocks of nodes will tend to group together big parts of the network even for small scales.

A notable approach, known as *geometric renormalization* [65, 59, 61, 146], is based on the hypothesis that the network's nodes can be naturally embedded in a metric space, whose geometry provides a natural way to identify groups of nearby nodes, reminiscent of spin blocks in the traditional real-space RG process [147]. These groups can then be collapsed into "super-nodes," providing a coarse-grained network description. However, this perspective encounters a fundamental limitation: networks are inherently topological structures, devoid of geometry, and thus need a topological notion of RG [60, 148].

Diffusion provides such a notion: it is a dynamical process that one can define on any combinatorial structure depending only on its topology. Diffusion on graphs, specifically, describes the dynamics of information flowing from node to node through edges, eventually becoming uniformly distributed on its connected components. This process is formalized as a first-order system of linear differential equations specified by the graph Laplacian matrix [1] L. One can then see the diffusion time as a *resolution parameter*: at short times, information only diffuses to neighboring nodes, revealing the local structure; at longer times, diffusion reaches nodes further apart, revealing the global structure. A recent proposal, the Laplacian renormalization group (LRG) [63, 68], leverages this observation to produce coarser descriptions of a network's structure by identifying groups of nodes that are strongly linked by diffusion at a given scale [149]. Moreover, one can adopt this same approach to define an *informational* notion of *scale-invariance*, based on the properties of the diffusion process via the Laplacian spectrum [150, 151], and thus different from the canonical concept of scale-freeness, which instead depends only on the degree distribution. Both of these results hinge on the formalism of network density matrices [152, 153, 66] to describe the complete behavior of information diffusion at a given scale.

Networks, however, are only part of the story. Great attention has recently been devoted to the study of higher-order networks: networks that encode multi-node interactions, going beyond the pairwise interactions of traditional networks [26, 34, 27]. However, little work still exists on RG approaches to higher-order networks. A direct generalization of the Laplacian RG approach [63], based on the multiorder Laplacian [37], was recently proposed [154]. This proposal is, however, *node-centric*: it only considers the diffusion of information from node to node. A parallel research line [150, 151, 94] focused on specific families of simplicial complexes and used renormalization group techniques to compute some notable statistical properties.

In this chapter, we follow Ref. [67] and propose a general renormalization group scheme for arbitrary higher-order networks. Our approach uses a higher-order notion of diffusion that we formally define by introducing the *cross-order Laplacian*. In this new diffusion process, information can flow through hyperedges of any order k via hyperedges of any other order m. This proposal provides a natural generalization of previous ones [66, 154], which are restricted to node-node diffusion. By studying the properties of this diffusion via the cross-order Laplacians, our approach allows us to probe the existence of characteristic scales, or—crucially—their absence (*scale-invariance*) in higher-order networks at each order. In

[1] In the language of simplicial complexes $L = L^0$, where L^0 is the order zero combinatorial Hodge Laplacian (Definition 0.19). See Remark 0.4

particular, we first define the appropriate Laplacian matrices to describe generic higher-order diffusion. We then leverage them to define *(i)* a higher-order notion of informational scale-invariance through the von Neumann entropy and entropic susceptibility, and *(ii)* an explicit RG scheme informed by a chosen higher-order diffusion process. Using these tools, we extract a cross-order scale signature in higher-order networks obtained from synthetic models and real-world data and show that in most cases, scale-invariance is found only under the lens of specific orders, suggesting the existence of underlying order-specific processes.

2.2 Higher-order networks and their structure

Higher-order networks. In general, when talking about higher-order networks, we mean a mathematical structure that generalizes the notion of a graph by introducing multi- (or higher-) order interactions between its nodes. Simplicial complexes (Definition 0.1) capture this notion, by including triangles, tetrahedra and their higher-dimensional analogues along with nodes and edges. However, the definition of abstract simplicial complex that we employed in this thesis is quite restrictive. This is due to the fact that Definition 0.1 requires the family of simplices to be *closed by inclusion*, i.e. any subset of simplex is also a simplex. This assumption can be too strong for some classes of relevant higher-order structures. The classical example is given by social interaction networks: a group of three friends can exist without necessarily the three pairwise friendships between its members.

Therefore, in this chapter we move away from simplicial complexes and deal with a more flexible, less structured kind of combinatorial object called *hypergraph*.

Definition 2.1 (Hypergraph)**.**

Let V be a finite set of vertices. A **hypergraph** X is a set of nonempty subsets of V.

It is clear from Definition 2.1 that an abstract simplicial complex is a hypergraph with the additional property of being closed under inclusion.

Let X be a hypergraph. Any element of X is called **hyperedge** and its **order** is defined as its cardinality reduced by one. Hyperedges in a hypergraph take the same role of simplices in a simplicial complex and thus we denote them in an analogous way with $\sigma \in X$ A vertex will thus be a 0-hyperedge •, an edge a 1-hyperedge  , a triangle a 2-hyperedge  , and

so on. If $\eta \in X$ is a subset of a k-hyperedge σ, then we say that η is a **face** of σ. We write X_k to denote the set of k-hyperedges of X and call n_k its cardinality.

Having fixed an order $k \in \mathbb{N}$, we want to study the scales of X from the perspective of its k-hyperedges. This approach, note, is in line with the simplicial Kuramoto model of Chapter 1, as we are not merely adding higher-order interactions to influence the node dynamics but we are *shifting the point of view* from the nodes to the interactions themselves.

In Ref. [63], it is argued that a diffusion process can be seen as a "telescopic scanner" of a (pairwise) network capable of extracting multiscale information about its structure. This fact suggests that we may study our hypergraph's k-th order properties through a diffusion process that takes place on the k-hyperedges.

Higher-order diffusion. A diffusion-like process on the hyperedges of a higher-order network can be defined in multiple ways, each associated with a different Laplacian matrix. In the case of simplicial complexes, it is natural natural to consider the k-th order combinatorial Hodge Laplacian (Definition 0.19), and it is in fact the one most commonly studied [155, 156], mainly for its topological significance (Theorem 0.4). A diffusion process on the k-hyperedges [52] can indeed be defined with the Hodge Laplacian, making it a natural candidate for our approach. However, we can easily see that interpreting the Hodge Laplacian as the generator of a "standard" diffusion process presents some physical criticalities.

The elements of the combinatorial Hodge Laplacian matrix Definition 0.19 can be written as

$$(L^k)_{ij} = \begin{cases} \deg_{(k,k+1)}(\sigma_i) + k + 1 & \text{if } i = j \\ 1 & \text{if } i \neq j,\ \sigma_i \cap \sigma_j \neq \emptyset, \forall \eta \in X\ \sigma_i \cup \sigma_j \not\subset \eta \text{ and } \sigma_i \sim \sigma_j \\ -1 & \text{if } i \neq j,\ \sigma_i \cap \sigma_j \neq \emptyset, \forall \eta \in X\ \sigma_i \cup \sigma_j \not\subset \eta \text{ and } \sigma_i \nsim \sigma_j \end{cases}, \tag{2.1}$$

where $\deg_{k,k+1}(\sigma_i)$ is the generalized higher-order *degree* of σ_i, i.e. the number of $(k+1)$-simplices that contain σ_i. The pattern of non-zero elements of Equation (2.1) tell us that L^k describes relations among k-simplices that are adjacent from below but are *not* adjacent from above. When two k-simplices σ, η satisfy this condition, they are said to be **parallel neighbors** [157] and we write $\sigma \parallel \eta$.

Moreover, there are other notable differences between this matrix and the standard graph Laplacian. First, the oriented nature of the simplices is such that the extra-diagonal elements

can be both -1 and $+1$, instead of being all -1. Second, the rows in general do not sum to zero. This last property comes from the fact that the vector $\mathbb{1} = (1, \dots, 1)^\top$ does not belong in general to $\ker L^k$, which instead contains the harmonic cochains corresponding to the k-th homology classes of X (Theorem 0.4).

These two properties make it so that diffusion through the Hodge-Laplacian is not a standard "physical" diffusion process. The fact that $\mathbb{1} \notin \ker L^k$, for instance, means that the total information present in the k-simplices is *not* conserved in time, as, under the linear diffusion dynamics $\dot{x}(\tau) = -L^k x(\tau)$

$$\mathbb{1} \notin \ker L^k \implies \frac{\mathrm{d}}{\mathrm{d}\tau} \sum_i x_i(\tau) = \frac{\mathrm{d}}{\mathrm{d}\tau} \mathbb{1}^\top x(\tau) = \mathbb{1}^\top \dot{x}(\tau) = -\mathbb{1}^\top L^k x(\tau) \neq 0. \tag{2.2}$$

Moreover, the oriented nature of the Hodge Laplacian manifested through the positive and negative extra diagonal terms means that another fundamental property of the graph Laplacian diffusion fails. If the initial configuration is non-negative $x(0) \geq 0$ it is not necessarily true that $x(\tau) \geq 0 \; \forall \tau \geq 0$, like it happens with the graph Laplacian. In other words, this means that a positive distribution of information diffusing through the network can result in the emergence of negative information values.

Cross-order Laplacians. Due to this lack of clear physical interpretability, it is difficult to directly employ the Hodge Laplacian in the Laplacian renormalization framework proposed in Ref. [66]. Moreover, the Hodge Laplacian cannot be naively extended to the general hypergraph setting.

Thus, we define a new family of Laplacian matrices that can describe a plethora of higher-order relations while maintaining a form analogous to the canonical graph Laplacian. We do this by taking inspiration from the general theory of combinatorial complexes [158] and the hypergraph Laplacian of Ref. [159].

Formally, we fix a number $k \in \mathbb{N}$, which we call **diffusion order**, to describe a diffusion process on the k-hyperedges of the hypergraph. We then need to decide how the process occurs, specifically, how hyperedges are "connected" so that information can flow between them. The most natural approach is to extract this information from the structure of X by employing a notion of *adjacency* among hyperedges. In general, two k-hyperedges σ and η can be adjacent in two ways:

- σ, η are *m-adjacent from above*, when there is an m-hyperedge ξ, with $m > k$, containing both of them $\xi \supseteq \sigma \cup \eta$;
- σ, η are *m-adjacent from below*, when they share a common m-face $\xi \subseteq \sigma \cap \eta$.

Definition 2.2 (Adjacency number)**.**

We define the **adjacency number** $a_{(k,m)}$ of a pair of *distinct* k-hyperedges $\sigma \neq \eta \in X$ as

$$a_{(k,m)}(\sigma, \eta) = \begin{cases} |\{\lambda \in X_m | \lambda \subseteq \sigma \cap \eta\}| \text{ if } m < k \\ |\{\lambda \in X_m | \sigma \cup \eta \subseteq \lambda\}| \text{ if } m > k \\ 0 \text{ if } m = k \end{cases} \tag{2.3}$$

so that we may consider η and σ to be m-adjacent $\sigma \overset{m}{\sim} \eta$ when $a_{(k,m)}(\sigma, \eta) > 0$.

Intuitively, $a_{(k,m)}(\sigma, \eta)$ counts the number of m-hyperedges connecting σ and η, allowing us to differentiate between different "strengths" of adjacency. For example, $a_{(2,0)} = 1$ when two triangles share a single vertex  and $a_{(2,0)} = 2$ when they share a pair of nodes  .

The (k, m)-adjacency relations can be formalized into different **adjacency matrices**, analogous to those defined in Refs. [158, 160]. If we index the k-hyperedges as $\sigma_1, \ldots, \sigma_{n_k}$, we can define the (k, m)-adjacency matrix with *diffusion order* k and *interaction order* m as the square $n_k \times n_k$ matrix with elements

$$(A_{(k,m)})_{ij} = a_{(k,m)}(\sigma_i, \sigma_j). \tag{2.4}$$

The adjacency matrix $A_{(1,2)}$, for example, describes how edges (1-hyperedges) are connected through triangles (2-hyperedges), while $A_{(3,0)}$ tells us how tetrahedra (3-hyperedges) are attached to one another through vertices (0-hyperedges). Notice how (k, m)-adjacency is equivalent to a weighted version of the q-nearness described in [161–163, 162] when X is a simplicial complex and we restrict our attention only to k-simplices.

The matrix $A_{(k,m)}$ can be seen as the adjacency matrix of a weighted undirected graph, which we call **adjacency graph** $G_{(k,m)}$ (see Figure 2.2), whose nodes are the k-hyperedges and the edges are their adjacency relations given by m-hyperedges. In particular, $G_{(0,1)}$ is the graph underlying the higher-order network (i.e. $X_0 \cup X_1$) and $G_{(1,0)}$ corresponds to its line

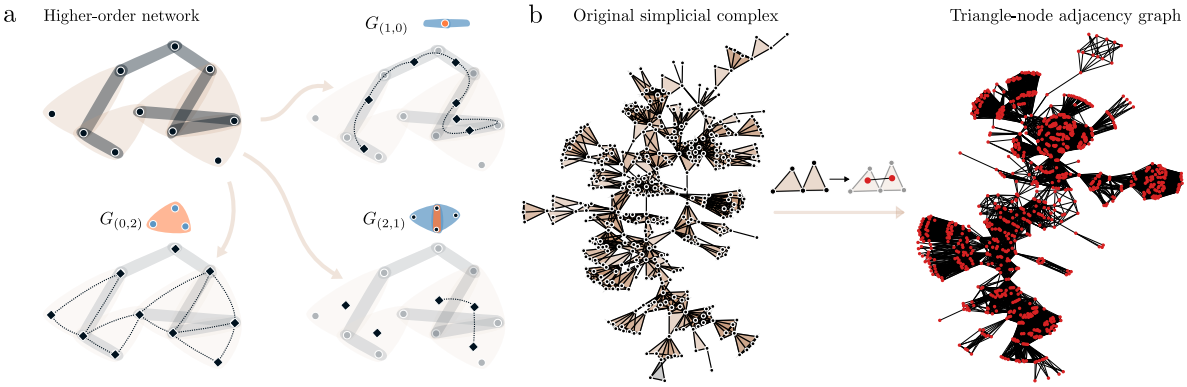

Fig. 2.2 **Adjacency grahps. a.** Schematic representation of how the adjacency graphs are built for a hypergraph. Notice how there can be cases like $G_{(0,2)}$, where the edges are weighted (here represented with multi-edges). **b.** A larger-scale simplicial complex and its $(2,0)$-adjacency graph.

graph [164], i.e. the graph where nodes correspond to edges and edges to nodes. To each weighted adjacency matrix, we can associate its (weighted) Laplacian.

Definition 2.3 (Cross-order Laplacian)**.**

The **cross-order Laplacian** of order (k,m) a hypergraph X is defined as the Laplacian matrix of the (k,m)-adjancency graph through the usual formula

$$L^{\times}_{(k,m)} = \mathrm{diag}(\deg_{(k,m)}) - A_{(k,m)}, \tag{2.5}$$

where $\deg_{(k,m)}$ is the vector containing the higher-order (k,m)-degrees, i.e. the row-sums of $A_{(k,m)}$

$$\deg_{(k,m)}(\sigma) = \sum_{\eta \in X_k} a_{(k,m)}(\sigma,\eta). \tag{2.6}$$

The matrix $L^{\times}_{(k,m)}$, being a weighted graph Laplacian, is symmetric, positive semidefinite and has an eigenvalue 0 with multiplicity given by the number of connected components of its underlying graph $G_{(k,m)}$. We note that our definition includes two existing families of Laplacians as particular cases:

(i) the "vertex-centric" higher-order Laplacians, describing how vertices can exchange information through hyperedges (e.g., the generalized Laplacians of Ref. [37]), as cross-order Laplacians of the form $L^{\times}_{(0,m)}$;

(ii) Hodge-like Laplacians, where diffusion happens between hyperedges through hyperedges of a directly adjacent order, as $L^{\times}_{(k,k\pm1)}$.

Cross-order Laplacians generalize these two notions, allowing the description of diffusion processes which, as with Hodge Laplacians, can happen on hyperedges of any order and, as with generalized Laplacians, can "jump" orders, connecting them with hyperedges of any other order.

Figure 2.2a clearly shows that the same hypergraph X can support vastly different dynamical structures, depending on the adjacency graph we use to probe it. There, for example, we see three cases:

- $G_{(1,0)}$ shows that information can diffuse only on a line.
- $G_{(0,2)}$ shows that information can spread freely through a large number of edges but, however, is prevented from reaching the top node due to the lack of triangles connecting it to the others.
- $G_{(2,1)}$ shows that information is severely restrained from the lack of edge connections between triangles. This leads to a sparse adjacency graph where some nodes (representing triangles) are isolated.

2.3 Statistical physics of higher-order diffusion processes

Cross-order diffusion. From now on, we will consider the high-order diffusion process on k-hyperedges through m-hyperedges on a given hypergraph X, and thus omit the (k, m) notation. Such a process can be easily written as the linear, first-order ODE

$$\dot{x}(\tau) = -L^{\times} x(\tau), \tag{2.7}$$

where $x(\tau) \in \mathbb{R}^{n_k}$ is a real scalar function on the k-hyperedges. Equation (2.7) can be solved with the time propagation operator or **heat kernel** at time $\tau > 0$

$$\rho(\tau) = e^{-\tau L^{\times}}, \tag{2.8}$$

so that

$$x(\tau) = \rho(\tau)x(0) \;\; \forall \tau \geq 0 \tag{2.9}$$

Due to linearity, we can think of the j-th column of $\rho(\tau)$ as describing the distribution of information over the k-hyperedges at time τ, when the total information is concentrated in a single k-hyperedge σ_j at time $\tau = 0$.

It turns out that we can derive interesting aggregate measures from the heat kernel, both for the standard Laplacian on networks [152] and for Hodge Laplacians on simplicial complexes [165]. These can help us probe the characteristic scales of the structure under consideration. First, ρ is normalized to a **density matrix** [152]

$$\hat{\rho}(\tau) = \frac{e^{-\tau L^{\times}}}{Z(\tau)} \tag{2.10}$$

where $Z(\tau) = \mathrm{Tr}(e^{-\tau L^{\times}})$ is called the **return probability**, describing how much of the information has remained "trapped" and did not diffuse at time τ [166].

The density matrix, which is characterized by $\mathrm{Tr}(\hat{\rho}(\tau)) = 1$ is a well-studied object that is used in quantum information theory [167] to describe the outcomes of measurements performed on quantum physical systems characterized by entanglement. For each density matrix, we define the following entropic measure, generalized from [152] to hypergraphs.

Definition 2.4 (von Neumann Entropy)**.**

Given the density matrix of Equation (2.10) $\hat{\rho}(\tau)$, we define its **von Neumann entropy** as

$$S(\tau) = -\mathrm{Tr}(\hat{\rho}(\tau) \log \hat{\rho}(\tau)).$$

Remark 2.1 (Spectral interpretation of the von Neumann entropy)**.** *The von Neumann entropy of our density matrix can be interpreted by taking a spectral point of view on the cross-order Laplacian. In fact, if λ_i is the i-th eigenvalue of $L^{\times}$, with $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{n_k}$, Definition 2.4 can be rewritten as*

$$S(\tau) = -\sum_{i=1}^{n_k} \frac{e^{-\tau\lambda_i}}{\sum_j e^{-\tau\lambda_j}} \log\left(\frac{e^{-\tau\lambda_i}}{\sum_j e^{-\tau\lambda_j}}\right), \tag{2.11}$$

which corresponds to the classical Shannon entropy *of the probability distribution p given by the normalized exponential of the eigenvalues $p_i = e^{-\tau\lambda_i}/Z(\tau)$.*

Moreover, if we move to the orthonormal basis $U = [U_1, \dots, U_{n_k}]$ of the eigenvectors of $L^{\times}$, the linear system of ODEs Equation (2.7) is diagonalized, becoming, if $u = U^{\top}x$

$$\dot{u}_i(\tau) = -\lambda_i u_i(\tau) \implies u_i(\tau) = e^{-\tau\lambda_i}.$$

Therefore, any solution to the diffusion equation can be written as a weighted sum of the graph eigenmodes with weights exponentially decreasing in time with rate given by the Laplacian eigenvalues

$$x(\tau) = \sum_{i=1}^{n_k} e^{-\tau\lambda_i} U_i.$$

The interpretation of the von Neumann entropy of Equation (2.11) (and thus Definition 2.4) becomes now clearer: $S(\tau)$ gives us the entropy of the distribution of eigenmodes that contribute to the process at time τ.

Proposition 2.1 (Limits of the von Neumann entropy)**.** *Let X be a hypergraph, $L^{\times}_{(k,m)}$ its (k,m) cross-order Laplacian, and $\beta_0^{(k,m)}$ the number of connected components of the adjacency graph $G_{(k,m)}$. Then*

$$\lim_{\tau\to 0} S(\tau) = \log n_k, \quad \lim_{\tau\to\infty} S(\tau) = \log \beta_0^{(k,m)}.$$

Proof. For $\tau = 0$, we have that $Z(\tau) = n_k$ and $e^{-\tau\lambda_i} = 1$, meaning that $S(0)$ is the entropy of the uniform distribution over n_k items, i.e. $S(0) = \log n_k$. For $\tau \to \infty$, all the terms associated with positive eigenvalues go to 0, $\lambda_i > 0 \implies e^{-\tau\lambda_i} \to 0$, leaving us only with the 0 eigenvalues. It follows from Theorem 0.4 that the number of zero eigenvalues coincides with the number of connected components of the underlying graph $G_{(k,m)}$ and thus $S(\tau)$ converges to the uniform distribution over $\beta_0^{(k,m)}$ items, $S(\tau) \to \log \beta_0^{(k,m)}$. □

Proposition 2.1 establishes S as an essential measure to track the renormalization of X as it can be interpreted as *the logarithm of the number of effective dynamical units* as the scale changes. In fact, it tells us that:

- For small diffusion times/small scales $\tau \approx 0$, the number of states is $e^{S(\tau)} \approx n_k$ i.e. all the k-hyperedges are independent units.
- For very large diffusion times/large scales $\tau \gg 0$, the number of states is $e^{S(\tau)} \approx \beta_0^{(k,m)}$. Here, we zoomed out to such a large scale that the only independent dynamical units are the connected components of k-hyperedges that do not communicate with each other.

- For intermediate scales, therefore, the k-hyperedges that behave in the same way are automatically merged in $S(\tau)$ and counted as the same dynamical unit.

Following the method proposed in Ref. [66], we study the scale behavior of X through the logarithmic derivative of the von Neumann entropy, known as the **entropic susceptibility**:

$$C(\tau) = -\frac{\mathrm{d}S}{\mathrm{d}\log\tau}. \tag{2.12}$$

Keeping in mind the interpretation we gave above, the entropic susceptibility thus measures the rate of variation of the effective (log of the) number of dynamical states or the rate of information acquired about the network during the diffusion dynamics at time τ [168].

These two measures, in the network case ($k = 0$, $m = 1$), have been fruitfully leveraged to extract key information about the network's structural organization [152, 153, 169, 170]. In particular, the maxima and minima of the entropic susceptibility, associated with times of fast deceleration and acceleration of the diffusion process, were shown to correspond to the presence of characteristic scales [66, 63]. Most importantly, we use $C(\tau)$ to define a notion of *scale-invariance* [63, 168] that tells us which networks can be renormalized.

Definition 2.5 (Informational scale-invariance)**.**

If the entropic susceptibility $C(\tau)$ of a hypergraph X is constant over a time range $I = [\tau_{\min}, \tau_{\max}]$, we say that X exhibits **informational scale-invariance** in the range of scales I.

Explicitly writing orders (k, m) again: it is known, for instance, that the (network, $k = 0$ and $m = 1$) entropic susceptibility $C_{(0,1)}(\tau)$ shows a large plateau in the case of grid graphs, Barabási-Albert networks, and random trees, all of which are examples of *self-similar* structures. Interestingly, entropic susceptibility and the associated notion of scale-invariance can be related to the concept of spectral dimension [171–173], which intuitively measures the dimensionality "perceived" by a diffusion process taking place on a manifold or, in our case, a combinatorial structure.

Informational scale-invariance and spectral dimension. To better understand the meaning of the definition of scale-invariance employed in this chapter and in Ref. [66], we can look at it

from two different angles. First, note that

$$C(\tau) = C^* \iff \frac{\mathrm{d}C(\tau)}{\mathrm{d}\log\tau} = 0 \iff \frac{\mathrm{d}^2 S(\tau)}{\mathrm{d}(\log\tau)^2} = 0 \ \ \forall \tau \in I$$

which means that scale-invariance is associated to a range of times where the rate of change of the entropy (in logarithmic scale) is constant, i.e. $S(\tau)$ does not accelerate nor decelerate.

From another point of view, it is interesting to look at the relation between the entropic susceptibility and the *spectral dimension* of the adjacency graphs of the simplicial complex. The spectral dimension [171–173] is a generalized notion of *dimensionality* that measures the dimension "perceived" by a diffusion process.

Definition 2.6 (Spectral dimension)**.**

We define the spectral dimension $D_s(\tau)$ as the derivative w.r.t. the logarithmic diffusion time of the logarithm of the return probability $Z(\tau)$

$$D_s(\tau) = -2\frac{\mathrm{d}\log Z(\tau)}{\mathrm{d}\log\tau}.$$

Here, we think of τ as a *scale* instead of a time parameter, meaning that $D_s(\tau)$ measures dimensionality at scale τ.

On d-dimensional flat manifolds $D_s(\tau) = d$ for all τ, while on d-dimensional (periodic) lattices $D_s(\tau)$ shows a large plateau whose value corresponds exactly to d [173]. In general, when $Z(\tau) \propto \tau^{-d/2}$ in a scale interval $I = [\tau_{\min}, \tau_{\max}]$, $D_s(\tau)$ has a plateau equal to d. This implies that, at these specific temporal resolutions, the space's spectral dimension observed through the diffusion process remains consistent. Such behavior suggests the existence of an inherent dimensionality within the space at these particular scales.

Note that Definition 2.6 generalizes to a scale-dependent quantity the more familiar notion of spectral dimension used for network [174], that is the fact that the density of eigenvalues of the Laplacian is $p(\lambda) \simeq \lambda^{d/2-1}$ for small λ. Indeed, if that is the case, we have that

$$Z(\tau) = \int_0^\infty e^{-\tau\lambda} p(\lambda)\mathrm{d}\lambda \approx \int_0^\infty e^{-\tau\lambda}\lambda^{d/2-1}\mathrm{d}\lambda = \tau^{-d/2}\Gamma(d/2),$$

and thus

$$D_s(\tau) = -2\frac{\mathrm{d}Z(\tau)}{\mathrm{d}\log\tau} \approx d.$$

One can see that there is a strong relation between the entropic susceptibility (Equation (2.12)) and the spectral dimension, i.e.

$$C(\tau) = -\frac{1}{2}\frac{\mathrm{d}D_s(\tau)}{\mathrm{d}\log\tau} + \frac{1}{2}D_s(\tau). \tag{2.13}$$

From Equation (2.13), we thus find that informational scale-invariance (Definition 2.5) is equivalent to

$$C(\tau) = C^* \iff D_s(\tau) = a\tau + 2C^* \tag{2.14}$$

i.e. the spectral dimension varies linearly w.r.t. the scale. This result tells us that informational scale-invariance corresponds to spaces with a well-defined intrinsic dimensionality ($a = 0$), but also the case in which the dimension varies linearly with the scale ($a \neq 0$).

This last case is particularly interesting because it reveals a characteristic of informationally scale-invariant geometries that goes beyond well-defined spectral dimension. Indeed, it is not hard to see that $C(\tau) = C^*$ if and only if

$$Z(\tau) \propto e^{-\frac{a}{2}\tau}\tau^{-C^*},$$

meaning that in an informationally scale-invariant space the return probability decays with a power-law, as in any flat Euclidean space, dampened by a negative exponential factor. Intuitively, this means that as the diffusing particle explores the space, it finds exponentially more room to fill at each step, therefore drastically reducing the probability of it coming back to the origin. This is typical of hyperbolic-like spaces, where the volume of a geodesic ball grows exponentially fast with the radius.

Example 2.1 (Torus graphs)**.** *Let us consider, as a simple example, the case of two-dimensional lattice graphs with periodic boundary conditions, thus topologically identifying a torus. We choose two examples that clarify the meaning of entropic susceptibility of Equation* (2.12)*. The first is a 64×64 node lattice graph, thus with a total of 4096 nodes, and the second is a 140×30 node lattice graph with 4200 nodes. The first thus represents a torus with major and minor radius of the same length, the second a torus with major radius* much larger *than the minor radius. Figure 2.3 shows the two tori, their von Neumann entropy $S(\tau)$ (left) and entropic susceptibility $C(\tau)$ (right). In accordance with Proposition 2.1, the entropy begins at $\tau = 0$ with value equal to $\log n_0 \approx 8.3$ and then decreases until it goes to 0 for large enough scales. Note that the entropy curves of the two tori coincide for small τ. This can be interpreted by looking at the entropic susceptibility plot on the right. After the first peak in C, representing the scale of the lattice discretization, we see a first plateau corresponding to*

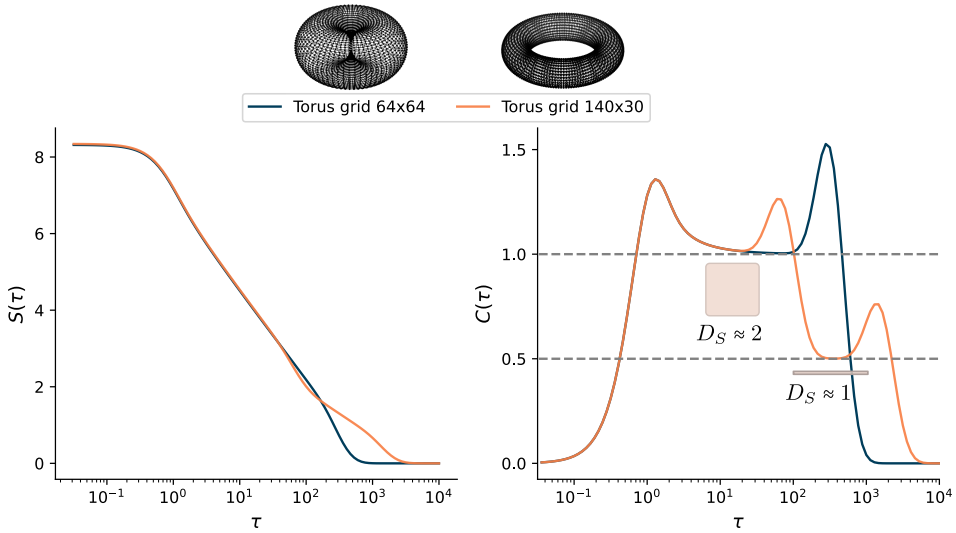

Fig. 2.3 Entropy $S(\tau)$ (left) and entropic susceptibility $C(\tau)$ (right) of two torus graphs of different ratios of major and minor radii. The plot on the right highlights the two prominent plateaus signifying scale-invariance.

$C^* \approx 1$. *Equation (2.14) thus tells us that, for this range of scales, both structures are identified as informationally scale-invariant with well-defined spectral dimension* $D_S \approx 2$. *This is due to the fact that both graphs are discretization of two-dimensional manifolds that are locally homeomorphic to* $\mathbb{R}^2$. *As the scale* τ *increases, the two tori display different behaviors. The entropic susceptibility of the 64×64 torus displays a final peak, corresponding to the scale of the whole graph, and then quickly falls to 0. The entropic susceptibility of the 140×30 torus, instead, has a second long plateau (we remind the reader that the x-axis is in logarithmic scale) with value* $C^* \approx \frac{1}{2}$. *This corresponds to the scale in which diffusion has become constant over the smaller radius, but not over the larger one, and thus the dynamical process sees the torus as a 1-dimensional manifold (*$D_S = 2C^* \approx 1$*), akin to a circle.*

Measuring scale-invariance. Following the considerations above, we want to quantify whether a higher-order network exhibits scale-invariance at order k via order m. To do so, we define the *scale-invariance parameter* (SIP) $P_{(k,m)}(\epsilon)$, as the logarithmic lifespan of the longest connected plateau of $C_{(k,m)}(\tau)$ w.r.t. a given tolerance $\epsilon > 0$ on its "flatness". In detail, given a value $y > 0$, we define the set

$$E_{(k,m)}(y;\epsilon) = \left\{\tau > 0 : |\log C_{(k,m)}(\tau) - y| < \epsilon\right\} \tag{2.15}$$

which, being the inverse image of the open interval $(y-\epsilon, y+\epsilon)$, will be given by a countable, disjoint union of open intervals (a_i, b_i)

$$E_{(k,m)}(y;\epsilon) = \coprod_i (a_i, b_i). \tag{2.16}$$

We then define the *scale-invariance parameter* as

$$P_{(k,m)}(\epsilon) = \max_{y>0} \max_i (\log b_i - \log a_i). \tag{2.17}$$

If $P_{(k,m)}(\epsilon)$ is large, we can say that the hypergraph is scale-invariant at order k via order m, while if it is close to 0 then $C_{(k,m)}(\tau)$ is peaked, thus signaling the presence of characteristic scales. Unless stated otherwise, in the numerical experiments we fix $\epsilon = 0.2$ and omit ϵ in the notation: $P_{(k,m)}(\epsilon) \equiv P_{(k,m)}$.

2.4 Cross-order Laplacian renormalization scheme

Let us assume that a hypergraph X has been recognized as scale-invariant with respect to $C_{(k,m)}(\tau)$ (Equation (2.12)). We would now like an algorithmic method to reduce it to a smaller, equivalent one that can still be recognized as scale-invariant. Similarly to the case of networks, we devise a *higher-order* Laplacian renormalization scheme consisting of the following steps:

1. Compute the cross-order Laplacian matrix $L^{\times}_{(k,m)}$.
2. Choose a diffusion time $\tau^* > 0$ corresponding to the scale at which to "zoom out".
3. Compute a partition of the k-hyperedges of X from the elements of the density matrix $\rho_{(k,m)}(\tau^*) \propto \exp(-\tau^* L^{\times}_{(k,m)})$, such that hyperedges in the same set are strongly linked by the diffusion process at time τ^*.
4. Coarse-grain X by merging its vertices according to the partition, to obtain a new, smaller X'.

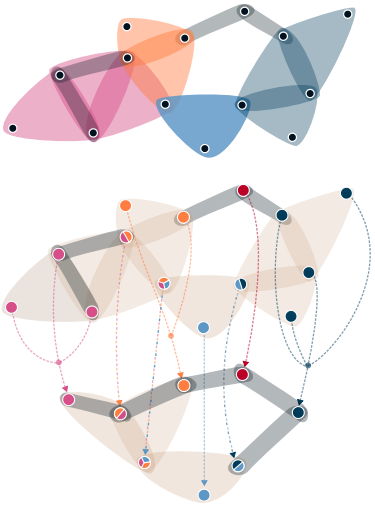

Fig. 2.4 Pictorial representation of our higher-order coarse-graining scheme with $k = 2$: *(i)* a partition of the 2-hyperedges, here represented with color, is obtained from $\rho(\tau)$, *(ii)* each vertex inherits a signature containing the labels of all 2-hyperedges it belongs to, *(iii)* vertices with the same signature are glued together, and hyperedges are induced from the starting hypergraph.

The process is then repeated, resulting in a sequence of hypergraphs with a decreasing or constant number of vertices. We name such a sequence **renormalization flow**.

In detail, staying faithful to the real-space Laplacian renormalization scheme, in step 3 we say that k-hyperedges σ and η are to be merged if the information which has flowed from σ to η is greater than the information which has stayed in σ or η. In practice, we build an auxiliary matrix ζ with

$$\zeta_{ij} = \begin{cases} 1 & \text{if } \rho_{ij} \geq \min(\rho_{ii}, \rho_{jj}) \\ 0 & \text{otherwise} \end{cases} \tag{2.18}$$

which can be thought as the adjacency matrix of a *metagraph* where the nodes represent the k-hyperedges and the edges connect those to be identified w.r.t. the diffusion process. A

partition of the k-hyperedges is then obtained simply by taking the connected components of the metagraph.

As prescribed by step 4, we need a coarse graining of X which is informed by the partition obtained in step 3. When the diffusion takes place between the vertices of a network, the process is quite simple [63]. It is enough to glue together vertices belonging to the same set into *super-vertices*, and place a *super-edge* $\{A, B\}$ between two super-vertices A and B if there are two vertices $a \in A$ and $b \in B$ in the original network such that $\{a, b\} \in X_1$.

If we want to renormalize the higher-order network w.r.t. a process taking place on general k-hyperedges more care is needed, as it is not clear how to naturally collapse hyperedges belonging to the same set in the partition. It is important to highlight that we cannot just coarse grain the adjacency graph w.r.t. the partition, as that will result in a smaller graph where most of the relational information contained in the hypergraph has been lost. To approach the problem, we propose the following method which, taking inspiration from Ref. [175], aims to reduce the problem to a node coarse-graining one.

Let us suppose that each k-hyperedge η has been assigned a label $\Lambda(\eta)$ specifying which set of the partition it belongs to. First, each k-hyperedge passes its label down to its vertices, or, from the other point of view, each vertex inherits the labels from all the k-hyperedges to which it belongs. Thus, we associate to each vertex v a set, called *signature* $\Sigma(v)$,

$$\Sigma(v) = \{\Lambda(\eta) \mid \eta \in X_k,\ v \in \eta\}\,, \tag{2.19}$$

which contains all the labels inherited. The coarse graining is then performed simply by identifying vertices possessing the same signature and then inducing higher-order hyperedges as explained above. If $\Sigma(v) = \emptyset$, meaning that i is not contained in any k-hyperedge, then we give it its own signature, so that it will be left unchanged by the transformation.

In other terms, we think of the signature as a labelling of the vertices of the hypergraph, i.e. a surjective map $f : X_0 \to S$ from the set of vertices to the set of signatures S. As it is explained in Ref. [176] for simplicial complexes, starting from this labelling, we can build another hypergraph X' whose vertices are the signatures, and whose hyperedges consist of the maps of the hyperedges of X through f, i.e.

$$X' = \{\{f(v_0), \dots, f(v_k)\} \mid \{v_0, \dots, v_k\} \in X_k,\ k \in \mathbb{N}\}\,. \tag{2.20}$$

Intuitively, step 3 identifies groups of k-hyperedges, which can be seen as generalizations of the spin blocks of Kadanoff's renormalization scheme [147]. In general, however, they will not be identical blocks of vertices, but their shape will reflect the structure of X (see Figure 2.4i). Once the partition of the k-hyperedges is obtained, we perform a coarse-graining step to aggregate the hyperedges belonging to the same block.

Note that the choice of the time τ^* heavily influences the renormalization process. If τ^* is too low, no vertices will be merged, whereas if τ^* grows large enough, every family of connected k-hyperedges in $G_{(k,m)}$ will get the same label ($S(\tau^*) = \log(\beta_0^{(k,m)})$ as in Proposition 2.1).

2.5 Examples of application

We now show explicit examples of applications of this framework. We first focus on synthetic models of simplicial complexes to confirm that, in controlled situations, the cross-order renormalization group recovers exactly the scale-invariant structure and order of the underlying system. After this confirmation, we extract the *cross-order scale signatures* of some real-world datasets.

Pseudofractal simplicial complexes. The family of scale-free pseudofractal simplicial complexes [177] provides a simple example of self-similar higher-order networks. Simplicial complexes in this family are built starting with a single k-simplex and by iteratively attaching a k-simplex to each $(k-1)$-simplex already present in the complex (Figure 2.5a). We expect the evident hierarchical nature of these objects to be visible in the entropic susceptibility curve of one or more diffusion processes defined on it.

Figure 2.5b illustrates how the different entropic susceptibilities $C_{(k,m)}$ display different, non-trivial behaviors. Most importantly, the curves associated with $L^{\times}_{(1,2)}$ and $L^{\times}_{(2,1)}$ show oscillating plateaus that span multiple orders of magnitude in τ. These correspond to the different, well-separated scales resulting from the iterative construction process, which can be interpreted as an approximate form of higher-order informational scale-invariance. Surprisingly, even if the network underlying the simplicial complex is scale-free [178], the curve of $L^{\times}_{(0,1)}$ (the canonical graph Laplacian) is peaked, meaning that the self-similarity of the structure is not visible at the vertex-edge level.

These results are consistently observed when we increase the number of steps with which the simplicial complex is built. We find that almost all the scale-invariance parameters $P_{(k,m)}$

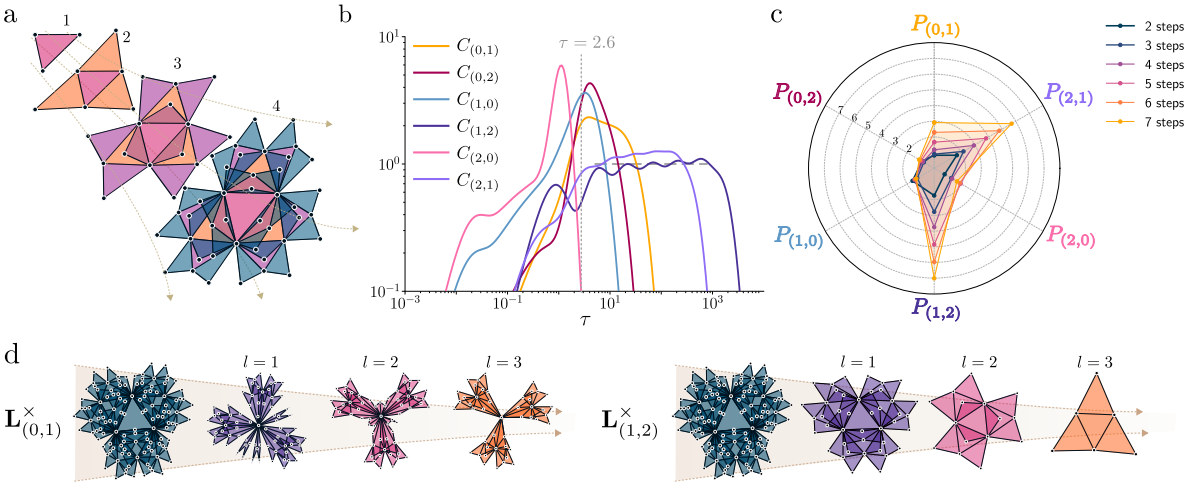

Fig. 2.5 **Scale-invariance and renormalization in pseudofractal simplicial complexes. a.** Graphical depiction of the first three steps of constructing the pseudofractal simplicial complex of dimension 2. **b.** Entropic susceptibilities of all the cross-order Laplacian matrices, computed for the 2-dimensional pseudofractal simplicial complex built with six steps (1095 vertices). **c.** Scale-invariance parameters varying with the number of steps with which the simplicial complex is built. **d.** On the left, the first three steps of the $L^{\times}_{(0,1)}$ renormalization scheme with $\tau^* = 0.2$. On the right, the first three steps of the $L^{\times}_{(1,2)}$ renormalization with $\tau^* = 2.6$.

increase with the number of construction iterations (Figure 2.5c), with the highest values being consistently obtained by $P_{(1,2)}$ and $P_{(2,1)}$, which correspond to the cross-order Laplacians naturally associated with the complex's growth (addition of triangles along edges). Moreover, the values of $P_{(2,0)}$, $P_{(1,0)}$ and $P_{(0,2)}$ remain approximately constant, while $P_{(0,1)}$ slowly increases, indicating the presence of a growing plateau (invisible in Figure 2.5b, where we show results for the realization with six steps). This observed behavior is also consistent with previous results [150] on the spectrum of $L^{\times}_{(0,1)}$ of pseudofractal simplicial complexes, which has power-law behavior for small eigenvalues when the number of vertices is large, i.e. a plateau in the entropic susceptibility (see Methods of Ref. [66]). However, the scale-invariant behavior is much more easily and quickly detectable when considering the entropic susceptibility of Laplacians associated with the complex's intrinsic growth process (between edges and triangles), even when the number of vertices is still small.

We confirm this observation by applying the renormalization scheme to a pseudofractal simplicial complex of dimension 2 and visualizing its evolution at each step. The renormalization based on $L^{\times}_{(1,2)}$ preserves the structure of the simplicial complex, perfectly reversing its construction process (Figure 2.5d, right). In contrast, the renormalization flow based on $L^{\times}_{(0,1)}$ destroys the pseudofractal structure and rapidly collapses the central vertices, resulting in a

"star-shaped" simplicial complex with no apparent relation to the original one (Figure 2.5d, left).

Network Geometry with flavor. While the higher-order scale-invariance for pseudofractal simplicial complexes is a relevant consistency check, this result is not particularly striking due to their evident hierarchical nature. We show here that analogous results can be found in the case of heterogeneous simplicial complexes, similar to what was previously observed in random scale-free networks and random trees [63]. To this end, a natural candidate is the family of simplicial complexes given by the *Network Geometry with flavor* (NGF) model [179, 178].

The NGF model [178] in dimension d is able to generate both hyperbolic manifolds and *scale-free* growing simplicial complexes, by progressively attaching d-simplices to $(d-1)$-simplices in a stochastic manner biased by the *flavor* parameter s. The three possible ways in which the growth process is realized, named *flavors*, are the following:

- $s = -1$, which allows for at most 2 d-simplices attached to any $(d-1)$-simplex, resulting in d-dimensional simplicial manifolds;
- $s = 1$, which glues d-simplices to $(d-1)$-simplices by preferential attachment;
- $s = 0$, which presents intermediate properties between the two.

Let us briefly describe the growth process of the model. At time $t = 1$ the NGF simplicial complex $X^{(1)}$ is made by a single d-simplex. At each time step $t > 1$ a new d-simplex is created and attached to one of the $(d-1)$-simplices present in $X^{(t)}$, chosen with probability

$$\Pi_{\sigma}^{[s]} = \frac{1}{Z^{[s]}(t)} e^{-\beta\epsilon_{\sigma}}(1 + s n_{\sigma}), \tag{2.21}$$

for each $(d-1)$-simplex σ, where

- β is the inverse temperature, which controls the amount of randomness in the process;
- ϵ_{σ} is the *energy* of the simplex σ, defined as the sum of the energies of its vertices, which in turn are sampled from a distribution $g(\epsilon)$;
- n_{σ} is the number of d-simplices which contain σ minus one;
- $Z^{[s]}(t)$ is the normalization constant.

From this it is easy to see that, when $s = -1$, all the $(d-1)$-simplices which are contained in exactly 2 d-simplices will have probability 0 to have another simplex attached to them. This results in $X^{(t)}$ being a simplicial manifold (with boundary) for every $t \geq 1$.

Due to its preferential attachment mechanism, we focus on the case $s = 1$, which leads to simplicial complexes with power-law distributed higher-order degree $\deg_{(d,d-1)}$, and, to manage computational complexity, we calculate entropic susceptibility curves only for relations predicted to be scale-free, alongside the vertex-edge $(0, 1)$ relation for comparison.

Although the NGF construction process is similar to the pseudofractals, the randomness does not allow for a clear separation of the hierarchical scales. The averaged curves of $C_{(d,d-1)}$ exhibit a smooth plateau anticipated by a small peak, which grows with higher dimensions, indicating the presence of a distinct microscopic scale. This is because the $(d-1, d)$-adjacency graph (of both d-dimensional NGFs and pseudofractals) is composed by $(d+1)$-cliques arranged in a tree-like structure, as d-simplices have $d+1$ $(d-1)$-faces. The first peak thus corresponds to the microscale associated with these cliques. Afterward, a scale-invariant regime ensues, illustrated by the near-perfect plateau of $C_{(d-1,d)}$. However, despite being associated with power-law degree distributions, other curves lack plateaus, showcasing non-trivial structural organization at specific scales.

In Figure 2.6b, a renormalization based on a higher-order relation ($L^{\times}_{(1,2)}$) shows better preservation of the structure of a 2-dimensional NGF simplicial complex with respect to the standard vertex-edge Laplacian renormalization. Notably, the $L^{\times}_{(0,1)}$-renormalization drastically reduces the number of vertices, collapsing them into a single super-vertex at the first peak of $C_{(0,1)}$ (Figure 2.6c, left). Conversely, $L^{\times}_{(1,2)}$-renormalization compresses the simplicial complex more gradually, revealing a clear transition point just before the first peak (Figure 2.6c, right). This distinction is further emphasized by tracking the evolution of the entropic susceptibility after one step of the renormalization flow in Figure 2.6d, where the plateau of $C_{(1,2)}$ is preserved by $L^{\times}_{(1,2)}$ but destroyed by $L^{\times}_{(0,1)}$.

Higher-order scale-invariance in real data. We can leverage scale-invariance parameters to provide effective higher-order signatures describing their hierarchical nature in a multifaceted way. Note that in general, unlike the regular cases discussed above, real networks are not the result of a simple construction process, and thus there is no clear way to identify a priori the pair (or pairs) (k, m) that can exhibit scale-invariance. Moreover, in a real-world context we cannot expect the presence of scale-invariance to be as clear-cut, with well-defined plateaus

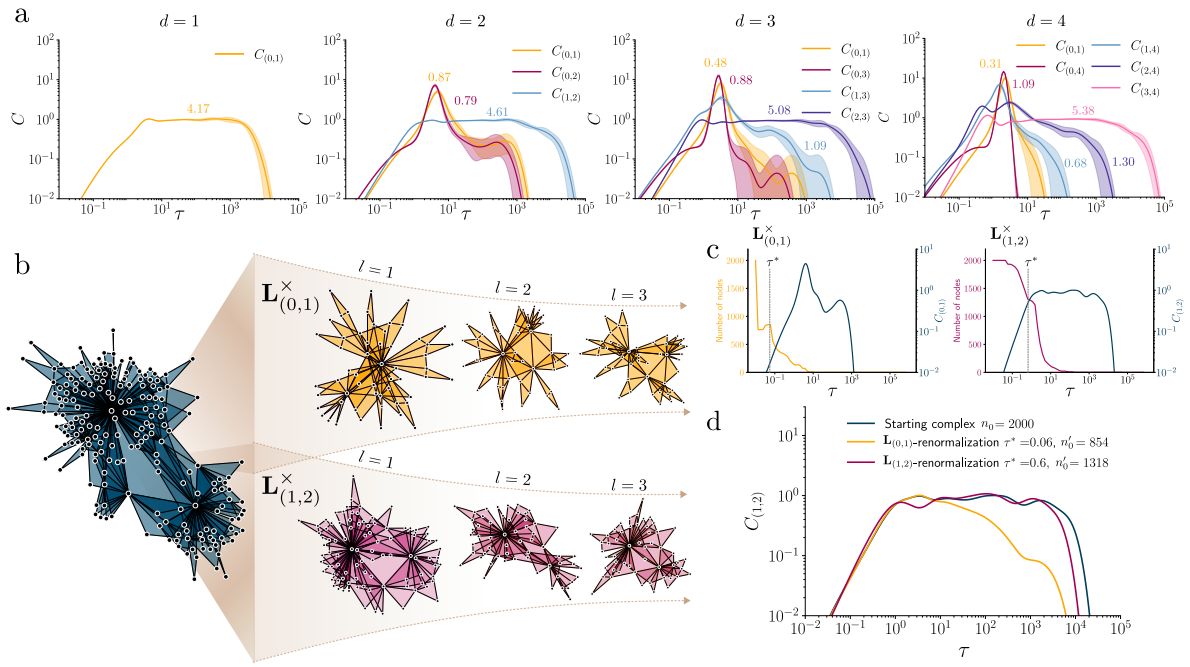

Fig. 2.6 **Higher-order Laplacian renormalization scheme applied to NGF simplicial complexes** **a.** Entropic susceptibilities, together with their 95% CIs over 10 repetitions, of the NGF simplicial complexes of flavor $s = 1$, dimensions $d \in \{1, 2, 3, 4\}$, $\beta = 5$ and 3000 vertices. The numbers display the scale-invariance parameters. **b.** A small 2-dimensional NGF simplicial complex is renormalized using $L^{\times}_{(0,1)}$ (top) and $L^{\times}_{(1,2)}$ (bottom). **c.** Left: $C_{(0,1)}$ together with the number of vertices of the complex after one step of $L^{\times}_{(0,1)}$-renormalization as a function of τ. Right: $C_{(1,2)}$ together with the number of vertices of the complex after one step of $L^{\times}_{(1,2)}$-renormalization as a function of τ. **d.** Evolution of the entropic susceptibility $C_{(1,2)}$ after the first step of the two types of renormalization considered. The starting NGF simplicial complex has 2000 vertices and is reduced to 854 vertices by $L^{\times}_{(0,1)}$ ($\tau^* = 0.06$) and to 1318 by $L^{\times}_{(1,2)}$ ($\tau^* = 0.6$).

as in the synthetic cases. In turn, fixing a maximum order $k_{\max}$, the most complete view is obtained by looking at all the $k_{\max}(k_{\max} - 1)$ entropic susceptibility curves up to that order.

We take real higher-order network datasets (Table 2.1) and consider their interactions up to the 4th order, i.e hyperedges of maximum 5 nodes. For each hypergraph, we compute all the $C_{(k,m)}$ with associated $P_{(k,m)}$ and build a null model by doing the same on their adjacency graphs randomized with a degree-preserving configuration model. In this way, we are able to obtain an effective signature of the hypergraphs encoding the amount of scale-invariance in each of its higher-order relations (Figure 2.7a).

Different behaviors emerge: the "flower-like" shape for the *High school contacts* [180] hypergraph, the "fan-like" shape —with higher values of scale-invariance in some specific orders— of the *Senate bills* [1], *House bills* [1], the *Diseasome* [181] and the *NDC Substances*

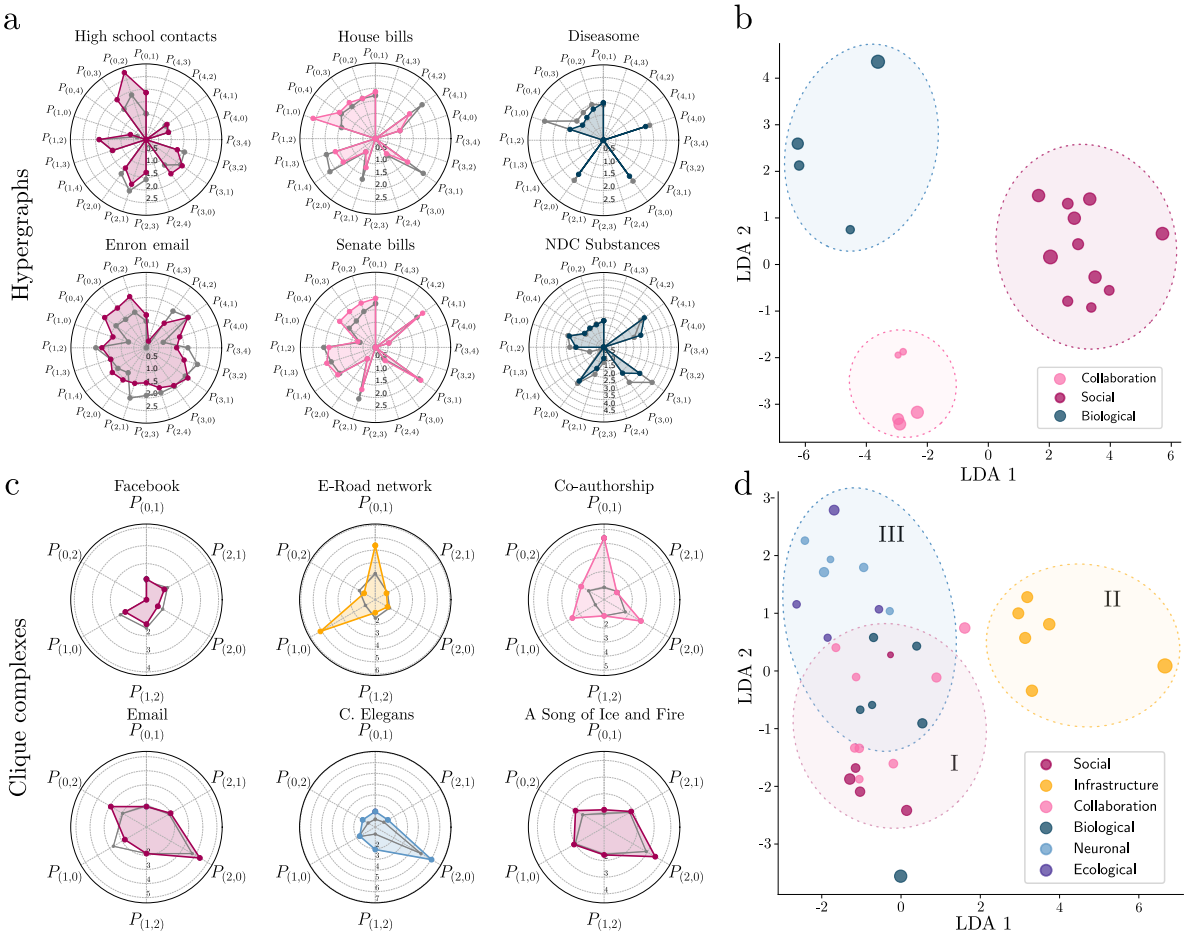

Fig. 2.7 **Higher-order scale-invariance in real data. a.** Scale-invariance parameters $P_{(k,m)}$ of six hypergraphs–truncated at order 4—obtained from real-world datasets. The gray lines and the associated regions represent the mean and 95% CI for the values of $P_{(k,m)}$ of the null model over 10 repetitions. **b.** LDA projection of a larger set of 20 real hypergraphs of different types, denoted by different colors, and with size proportional to the value of their highest SIP. The colored regions highlight clusters of nearby points that belong to the same class of hypergraphs. **c.** Scale-invariance parameters $P_{(k,m)}$ of six second-order clique complexes obtained from real-world network datasets. The gray lines and the associated regions represent the mean and 95% CI for the values of $P_{(k,m)}$ of the null model over 10 repetitions. **d.** LDA projection of 34 real clique complexes of different types, denoted by different colors, and with size proportional to the value of the network's highest SIP. The colored regions highlight clusters of nearby points that belong to the same class: social origin (I), infrastructural (II), biological origin (III).

[182] hypergraphs. Moreover, as for the *Enron Email* [182, 183], there can be a more uniform distribution of the SIPs values, without a clear dominance of orders. Note that, when the k-hyperedges are too sparse, information cannot flow through m-hyperedges, resulting in zero-valued SIPs for those specific orders.

Interestingly, even when we do not consider genuine higher-order networks but rather focus on networks, it is still possible to obtain meaningful information through our approach. We consider real pairwise network datasets (Table 2.2) and compute their associated *clique complexes* up to 2nd order, i.e. build simplicial complexes by considering their cliques of three nodes as 2-hyperedges. Notice that in this way we artificially lift a (pairwise) network to a higher-order one and, as a consequence, any resulting higher-order properties should be interpreted as describing the structural organization of cliques of nodes, and not of "genuine" higher-order interactions.

In Figure 2.7c we show the SIPs associated with some of these clique complexes, together with the null model. We find a variety of different behaviors, like the *E-road* network [184] and the network science *Co-authorship* network [185], which are strongly scale-invariant in the standard sense, i.e. in the relation between vertices and edges (high values of $P_{(0,1)}$ and $P_{(1,0)}$). Despite the inherent pairwise nature of the networks, we find a marked presence of high-order scale-invariance in some of them, like the University Rovira i Virgili *Email* network [186], the *C.Elegans* metabolic network [187–189] and the fictional *A Song of Ice and Fire* social network [190]. We stress again that this higher-order description is obtained without any assumptions on the presence of genuine higher-order interactions in these systems, but rather it is based only on nodes, edges, and the cliques they form. However, some aspects of the hierarchical scale-invariant structure of a network may be hidden from the node-centric point of view, but can be evident when looking at how simplices (in this case, cliques) are related to one another. This is in line with similar results on the persistent homology of weighted networks [191].

Finally, we use scale-invariance parameters as coordinates to embed the datasets in $\mathbb{R}^2$. In Figure 2.7b and Figure 2.7d, the linear projection–Linear Discriminant Analysis (LDA)[192]–of the resulting point cloud are showing heterogeneity in the amount of higher-order scale-invariance present in each dataset. Surprisingly, despite the coarseness of this measure, we see that the information contained in SIPs is enough to cluster the networks according to their nature, suggesting the presence of specific higher-order signatures associated with each type of data.

2.6 Discussion

By leveraging the Laplacian renormalization group, we developed a method to investigate the structure of higher-order networks via a renormalization procedure based on the connectivity

structure of higher-order relations, as encoded in the proposed cross-order Laplacian. We showed that the renormalization scheme can revert exactly the construction of higher-order structures that are self-similar by construction (the pseudofractal) and detect the correct order of the dominating growth mechanism in scale-invariant complexes with induced heterogeneous lower order structures (the NGF model). Armed with these results on controlled synthetic systems, we employed the entropic susceptibilities obtained in our scheme to build *scale-invariance profiles* for a set of real-world systems, revealing both different unexpected kinds of characteristic scales and scale-invariance at various orders and commonality of such profiles across systems belonging to the same domain.

From a technical point of view, the cross-order Laplacians bear both similarities and differences with previously defined higher-order Laplacians, notably the Hodge (or combinatorial) Laplacian [52] L^k, and the multi-order Laplacian $L^{(\mathrm{mul})}$ [37]. In fact, both the k-order Hodge Laplacian and cross-order Laplacians $L^{\times}_{(k,m)}$ (for any m) are defined on the hyperedges of order k, and thus are $n_k \times n_k$ matrices, where n_k is the number of k-hyperedges. However, the Hodge Laplacian is limited to adjacencies defined by boundary and coboundary relations, and thus, with $m = k \pm 1$, while the cross-order can capture arbitrary adjacencies through any m-hyperedges. Conversely, the multi-order Laplacian $L^{(\mathrm{mul})}$ is defined as the weighted sum of Laplacians defined on nodes and adjacent via hyperedges of any m order. It can be rewritten in our notation as $L^{(\mathrm{mul})} = \sum_{m=1}^{M} \omega_m L^{\times}_{(0,m)}$, where the ω_m are an arbitrary weighting scheme and M the maximal order considered. Similarly, we note that, while in this work we focused on renormalization based on specific pairs (k, m), important future work should understand the effect of extending the cross-order renormalization scheme to a *multicross*-order scheme, which could be achieved by considering combinations of hyperedges of dimension higher or lower than a certain threshold on the interaction order (e.g., $L_k^{(\mathrm{mul}\times)} = \sum_{m\in\{\mathbf{m}\}} \omega_m L^{\times}_{(k,m)}$ over a set $\{\mathbf{m}\}$). Also, while this work focused mainly on simplicial complexes for convenience and clearness of exposition, any combinatorial structure [158] built with a set of ranked sets is amenable to our scheme.

Finally, our results provide a new lens to address questions on the origin of different higher-order invariant structures in various domains [193, 194], on their effects on dynamical processes taking place on them [195, 195, 196] and their reducibility [197], as well as to the limits they might pose on the predictability and reconstruction of complex systems [198].

2.7 Datasets used

Real hypergraphs. We chose a heterogeneous family of 20 real hypergraphs from the repository hosted by the XGI package [1] and compute their SIPs up to the 4-th order. These hypergraph are of different types, from contact networks to gene-disease networks, and, as we did for clique-complexes, we divide them into three coarse families: "social", "collaboration" and "biological".

Name	**Type**	**Finer Type**	**Source**	n_0	n_1	n_2	n_3	n_4
High School Contacts	Social	Contacts	[180]	327	5498	2091	222	7
Primary School Contacts	Social	Contacts	[199–201]	242	7748	4600	347	9
Enron Email	Social	Email	[182, 183]	143	809	317	138	63
Eu Email	Social	Email	[182, 202, 203]	986	12753	4938	2294	1359
Lyon Hospital Contacts	Social	Contacts	[204, 201]	75	1107	657	58	2
Hypertext Contacts	Social	Contacts	[205, 201]	113	2103	302	18	7
InVS13	Social	Contacts	[206]	92	741	44	2	0
InVS15	Social	Contacts	[207]	217	4142	755	12	0
Malawi Village Contacts	Social	Contacts	[208]	84	341	86	4	0
Science Gallery	Social	Contacts	[205, 201]	410	2491	808	46	5
SFHH Conference	Social	Contacts	[207, 209, 199]	403	8268	1861	258	63
US Congress Bills	Collaboration	Bills	[182, 210, 211]	1718	13871	10156	7764	5780
House Bills	Collaboration	Bills	[1]	1494	7656	4369	3350	2607
House Committees	Collaboration	Committees	[1]	1290	1	2	11	25
Senate Bills	Collaboration	Bills	[1]	294	3280	3013	2394	1685
Senate Committees	Collaboration	Committees	[1]	282	0	0	9	22
Diseasome	Biological	Disease	[181]	516	153	92	26	25
DisGeNet	Biological	Disease	[212]	12368	157	139	93	66
Classes of NDC	Biological	Pharmaceutical	[182]	628	135	92	100	73
Substances of NDC	Biological	Pharmaceutical	[182]	3414	1027	717	516	498

Table 2.1 Datasets from XGI package [1] used for Figure 2.7a,b

Real networks. We then considered a larger set of pairwise networks taken from different network archives: *KONECT* [213], *ICON* [214] and *The Network Data Repository* [215]. The networks, whose details are found in Table 2.2, were chosen to belong to the following classes:

- *Infrastructure* — road networks and power grids;
- *Collaboration* — scientific collaboration networks;
- *Social* — email networks, real and fictional social networks;
- *Biological* — gene interaction networks and protein-protein interaction networks;

- *Connectome* — brain networks;
- *Ecological* — species food webs in different ecological environments.

Each one of the networks is lifted to a simplicial complex by computing the associated clique complex. This means taking all cliques up to a chosen order and seeing them as simplices. We then compute the scale-invariance parameters up to the second order, that is $P_{(0,1)}$, $P_{(0,2)}$, $P_{(1,0)}$, $P_{(1,2)}$, $P_{(2,0)}$ and $P_{(2,1)}$.

Name	**Type**	**Repository**	**Source**	n_0	n_1	n_2	n_3
Facebook	Social	KONECT	[216]	2890	2982	91	4
E-Road	Infrastructure	KONECT	[184]	1176	1418	32	0
Network Science Coauthorship	Collaboration	KONECT	[185]	1461	2742	3764	7159
Email	Social	KONECT	[186]	1133	5451	5343	3419
C. Elegans Metabolic	Biological	KONECT	[187–189]	453	2025	3284	2967
A Song of Ice and Fire	Social	KONECT	[190]	796	2823	5655	9331
Human Protein	Biological	KONECT	[217]	3133	6726	1047	142
Power Grid	Infrastructure	KONECT	[12]	4943	6595	651	90
Yeast Protein	Biological	KONECT	[218–221]	1872	2278	222	41
Hamsterer Friendships	Social	KONECT	[213]	1858	12534	16750	10015
German Highway System	Infrastructure	ICON	[222]	1168	1243	2	0
Cat Brain	Connectome	ICON	[223]	65	1139	3613	10125
Rhesus Brain	Connectome	ICON	[224]	91	582	1902	4048
Mouse Visual Cortex	Connectome	ICON	[225]	195	214	4	0
C. Elegans Posterior Nervous System	Connectome	ICON	[226]	272	4451	10005	16839
IUI Journal Authors	Collaboration	ICON	[227]	2288	3377	2411	1904
CoRA Citations	Collaboration	ICON	[228]	3585	11258	8631	3940
Flensburg Fjord Food Web	Ecological	ICON	[229]	180	1569	3440	4175
Hessen Traffic	Infrastructure	ICON	[230]	4660	6026	192	0
Minnesota Roads	Infrastructure	Net. Data. Rep.	[231]	2641	3302	53	0
Yeast Interactome	Biological	ICON	[232]	1278	1808	110	6
Barcelona Roads	Infrastructure	ICON	[233]	930	1798	164	1
Freshwater Stream Food Web	Ecological	ICON	[234]	113	832	124	0
Little Rock Lake Food Web	Ecological	ICON	[235]	183	2452	11292	35978
P. Pacificus Neural	Connectome	ICON	[236]	54	141	117	32
Slovenian Computer Scientists	Collaboration	ICON	[237]	262	596	488	224
New Zealand Scientific Collaboration	Collaboration	ICON	[238]	1511	4273	9337	21986
Computational Geometry	Collaboration	ICON	[239]	6158	11897	13587	20994
Physician Trust Network	Collaboration	ICON	[240]	243	924	627	220
Erdos Collaboration	Collaboration	ICON	[241]	5094	7514	1607	450
Facebook Food Pages	Social	Net. Data. Rep.	[242]	942	2101	0	0
Diseasome	Biological	Net. Data. Rep.	[181]	517	1189	1360	1391
C. Elegans Genes	Biological	Net. Data. Rep.	[243]	2989	4659	118	2
Florida Ecological	Ecological	Net. Data. Rep.	[244, 245]	128	2074	8436	14126

Table 2.2 Datasets used for Figure 2.7c,d

Topology & dynamics of neural network training

PART II

Training a neural network consists of navigating the complex geometry of a loss landscape to reach one of its deepest valleys. While gradient descent and its variants are the most widely used algorithms for this task, the standard visualization of optimization—as a trajectory in Euclidean parameter space rolling downhill—is somewhat misleading. Different parameter configurations can be *observationally equivalent*, encoding the exact same mathematical function [246]. These equivalences are not accidental; they are induced by the symmetries of the network architecture and activation functions, and they profoundly shape the geometry of the loss landscape.

As a consequence, minima are typically not isolated points but form high-dimensional manifolds with nontrivial geometry [247–249]. Moreover, the gradient and Hessian of the loss are constrained to obey specific structural laws [250, 251]. Optimization dynamics are therefore deeply intertwined with the symmetry-induced geometry of the parameter space, and gradient-based methods are inherently sensitive to these constraints [252, 253].

The success of deep learning has motivated extensive efforts to understand these geometric aspects. A prominent line of research focuses on the *neuromanifold*—the space of functions or probability distributions representable by a network—using the tools of information geometry [254]. While our work echoes the study of the neuromanifold by moving beyond flat Euclidean intuition, it is not equivalent. Rather than focusing on the space of functions, we investigate the *geometry of the optimization process itself*. Specifically, we study the structures that emerge from the interplay between symmetries and gradient-based dynamics, determining not just what a network can represent, but where the training process can actually lead.

In Chapters 3 and 4, we adopt a topological perspective on the constraints induced by symmetry groups on neural network optimization. Topology provides the natural tools to formalize global obstructions to optimization by studying properties preserved under continuous deformations. Our central objective is to characterize, in topological terms, when training trajectories are unable to freely explore parameter space to move from an arbitrary initialization to an optimal solution. This question is naturally phrased in terms of *connectedness*, and in particular through the zeroth *Betti number*, which counts the number of connected components of a space.

Our analysis is grounded in the observation that gradient flow training of networks with homogeneous activations gives rise to symmetry-induced conservation laws. These laws confine optimization trajectories to an algebraic variety—referred to as the *invariant set*—defined by a system of quadratic equations that depend on the network topology and the initialization. We study two fundamental pathologies of this invariant set: its possible disconnectedness and the

presence of singularities. Importantly, the existence of these topological obstructions does not depend on the specific loss function or training data; it is intrinsic to the interplay between geometry, topology, and symmetry in the parameter space.

Beyond their theoretical implications, these results shed light on the role of network topology in optimization and suggest principled approaches to model compression. We demonstrate how singular configurations correspond to disconnected subnetworks and how appropriately designed regularizers can drive training toward such singularities. This enables structure-agnostic pruning mechanisms grounded in the intrinsic geometry of the optimization space rather than heuristic sparsity constraints.

Related Work

Training dynamics of ReLU networks. A large body of work studies gradient-based optimization of neural networks with homogeneous activations. Convergence properties have been established for wide and mean-field regimes [255–257], and the implicit bias of these methods has been analyzed under various hypotheses, including initialization scale [258, 259] and overparameterization [260]. While these works primarily characterize the optimal solutions selected by the dynamics, our focus is on the global structure of the optimization space itself and situations where optimal solutions may be topologically unreachable from a given initialization. Closely related to our setting, [261] studies univariate two-layer ReLU classifiers, while [262] analyzes the analytical challenges arising from the non-smoothness of ReLU.

Symmetries and conserved quantities. The positive homogeneity of the ReLU activation implies that certain weight rescalings leave the network function unchanged, a property that underlies the critique of sharpness-based generalization measures [246]. Symmetry-induced conservation laws have been derived for gradient flow in both shallow and deep networks [251, 263, 264], enabling applications such as pruning [250] and explaining the automatic balancing of neuron weights [252]. While these laws are exact in gradient flow, they can be broken in finite step-size gradient descent, leading to modified trajectories and implicit regularization [265, 266].

Topology and geometry of the loss landscape. There is a rich history of investigating the geometry of loss landscapes to gain insight into training behavior. Much of this literature

focuses on the connectedness of loss sublevel sets [247, 267, 249], motivated by empirical findings that low-loss solutions can often be connected by simple curves [268, 269]. A complementary line of work analyzes the presence of spurious minima [270, 271]. Our work differs by focusing on the topology of the *training-invariant parameter space* (the invariant set) rather than the topology of the loss level sets themselves.

DAG architectures and pruning. General feedforward networks can be formalized as directed acyclic graphs (DAGs) [272, 273]. Such architectures naturally arise in biological systems and through pruning mechanisms like the lottery ticket hypothesis [274]. Pruning methods are generally divided into structured and unstructured approaches [275, 276]. Our contribution connects these methods to the geometry of the optimization space, showing how nuclear norm regularization promotes singular configurations that correspond to disconnected subnetworks.

Singularities and learning. Singularities are central to Singular Learning Theory [277] and the burgeoning field of neuro-algebraic geometry [278], which investigates the space of functions realizable by networks. The influence of singularities on optimization has been recognized since the early development of information geometry [279, 280]. Our work contributes to this lineage by identifying singularities intrinsic to the invariant set of gradient flow and analyzing their specific role in optimization and model compression.

Back to Table of Contents

Chapter 3

Topological obstruction to the training of shallow ReLU neural networks

Contents

3.1 Introduction

In this chapter, which is based on Ref. [264], we formalize the geometric constraints of the optimization process by investigating the training of shallow ReLU neural networks under gradient flow. While the preceding discussion established that symmetries shape the loss landscape, we demonstrate that these symmetries do more than just create equivalent minima; they define a rigid container for the entire optimization trajectory. We introduce and analyze the *invariant set*: an algebraic variety to which any gradient-based trajectory is strictly confined from initialization.

The central result of this chapter is the identification of *topological obstructions* that can fundamentally impede learning. By analytically computing the zeroth Betti number of the invariant set, we prove that for networks with a single scalar output, this space is not necessarily connected. This disconnectedness reveals a failure mode for neural network optimization: depending on the initialization, there exist regions of the parameter space that are fundamentally unreachable, regardless of the loss function or the number of training iterations. We show that these isolated components arise from "pathological" neurons whose weight signs are locked by symmetry-induced conservation laws, effectively partitioning the parameter space into mutually inaccessible zones.

Furthermore, we explore the interplay between these topological constraints and the discrete symmetries of the network, such as neuron permutation. We provide a refined counting of the "effectively" connected components, accounting for the fact that some topological gaps can be bridged by permutation symmetries while others remain absolute.

Main contributions. The main contributions of this chapter are summarized in the following.

1. We find that, for two-layer neural networks, the gradient flow trajectories lie on an invariant set, which can be factored as the Cartesian product of quadric hypersurfaces.
2. We analytically compute its singular Betti numbers (Definition 0.7).
3. We find that the invariant set can be disconnected when the network's output dimension is 1, leading to a clear topological obstruction.
4. We find that the obstruction is caused by "pathological" neurons that cannot change the sign of their output weight when trained with gradient flow.

5. We discuss the relation between the invariant set and the network's discrete symmetries, finding that if we consider permutations, the number of effective connected components scales linearly in the number of pathological neurons.

3.2 Setup and preliminaries

3.2.1 One-hidden layer neural network

Let us consider a two-layer neural network $f(\cdot;\theta) : \mathbb{R}^d \to \mathbb{R}^e$ specified by the function

$$f(x;\theta) = W^{(2)}\sigma(W^{(1)}x), \tag{3.1}$$

where $x \in \mathbb{R}^d$ is the input, $\theta = (W^{(1)}, W^{(2)})$ with $W^{(1)} \in \mathbb{R}^{l\times d}$ and $W^{(2)} \in \mathbb{R}^{e\times l}$ are the parameters, $\sigma : \mathbb{R} \to \mathbb{R}$ is the component-wise activation function and l is the number of neurons in the hidden layer. Notice that we consider a network with no biases, as it allows us a discussion with lighter notation. The case with biases is discussed in Section 3.10.

In this chapter, following previous work [252], we focus on a particular class of activation functions that includes the commonly used ReLU ($\sigma(z) = \max\{z, 0\}$) and Leaky ReLU ($\sigma(z) = \max\{z, \gamma\}$ with $0 \leq \gamma \leq 1$) activations.

Assumption 3.1. *We assume that the activation function σ is **positively homogeneous** of degree one, i.e. $\sigma(ax) = a\sigma(x)$ for every $a > 0$ and $x \in \mathbb{R}$.*

We note that Euler's homogeneous function theorem states that a positively homogeneous function satisfies $\sigma(x) = \sigma'(x)x$ for every x where σ is differentiable.

We call **parameter space** the Hilbert space of pairs of matrices

$$\Theta = \left\{\theta = (W^{(1)}, W^{(2)}) \mid W^{(1)} \in \mathbb{R}^{l\times d}, W^{(2)} \in \mathbb{R}^{e\times l}\right\},$$

equipped with the standard matrix sum, scalar multiplication operations, and the Frobenius inner product $\left\langle (W^{(1)}, W^{(2)}), (U^{(1)}, U^{(2)})\right\rangle = \sum_{k=1}^{2}\sum_{i,j} W_{ij}^{(k)}U_{ij}^{(k)}$. Any vector $\theta \in \Theta$ specifies the parameters of a neural network $f(x;\theta)$.

While elements of Θ contain all the network's parameters, in the following it will be necessary to examine the single hidden neurons and their associated parameters. The following

essential results tells us that for two-layer neural architectures, the parameter space nicely decomposes into the sum of its neurons' parameters.

Proposition 3.1. *For the two-layer neural network defined in Equation (3.1). Let* $k \in \{1, \dots, l\}$, $(e_{11}, e_{12}, \dots, e_{ll})$ *be the canonical basis of* $\mathbb{R}^{l\times l}$ *and*

$$\Theta_k = \left\{\theta_k = (e_{kk}W^{(1)}, W^{(2)}e_{kk}) \mid (W^{(1)}, W^{(2)}) \in \Theta\right\} \subset \Theta,$$

then it holds that $\Theta = \Theta_1 \oplus \cdots \oplus \Theta_l$.

Therefore, we can consider Θ_k as the **parameter space of the k-th hidden neuron**, which consists of the inputs and output weights of neuron k, namely the k-th row of $W^{(1)}$ and k-th column of $W^{(2)}$, respectively. For simplicity, when we work in Θ_k, we write $W_k^{(1)} := e_{kk}W^{(1)}$ and $W_k^{(2)} := W^{(2)}e_{kk}$. Interestingly, the decomposition of Proposition 3.1 only holds for two-layer neural networks and will be crucial to the formulations of this chapter's results.

Proof. Θ_k, notice, is the subspace of Θ consisting of the weight matrices $W^{(1)}$ with null rows except for the k-th one, and weight matrices $W^{(2)}$ with null columns except for the k-th one. We can check that $\Theta = \Theta_1 \oplus \cdots \oplus \Theta_l$, because, if I_l is the $l \times l$ identity matrix, $\sum e_{kk} = I_l$ so that

$$(W^{(1)}, W^{(2)}) = \sum_k (W_k^{(1)}, W_k^{(2)}) = \sum_k (e_{kk}W^{(1)}, W^{(2)}e_{kk}).$$

Moreover, $\Theta \cong \Theta_1 \times \cdots \times \Theta_l$ via the linear isomorphism

$$\theta = (W^{(1)}, W^{(2)}) \leftrightarrow (\theta_k)_{k=1}^l = \left((e_{kk}W^{(1)}, W^{(2)}e_{kk})\right)_{k=1}^l.$$

Orthogonality follows from the fact that $W_k^{(1)}$ has nonzero elements only in the k-th row, meaning that, the inner product with $W_j^{(1)}$, whose nonzero elements are contained in the j-th column, will be 0 whenever $j \neq k$. This, we see, is equivalent to decomposing the neural network of Equation (3.1) into the computations of the single hidden neurons. Indeed, let $f(x; \theta_k) := f(x; (e_{kk}W^{(1)}, W^{(2)}e_{kk}))$, then, considering that $e_{kk}e_{kk} = e_{kk}$ and that $\sigma(e_{kk}v) = e_{kk}\sigma(v)\ \forall v \in \mathbb{R}$, it holds that

$$f(x; \theta_k) = W^{(2)}e_{kk}\sigma(W^{(1)}x)\ \forall k = 1, \dots, l.$$

Therefore $\sum_k f(x; \theta_k) = f(x; \theta)$. □

3.2.2 Symmetries and observationally equivalent networks

It is well known that the geometry of the parameter space Θ is heavily influenced by the properties of the activation function. Especially, the commutativity of the activation function with some classes of transformations can result in the latter having no effect on the function implemented by the neural network. This means that, in general, the mapping from the parameter space to the hypothesis class of functions is *not injective*. This means that two-layer neural networks (and neural networks) belong to the class of so-called *non-identifiable* models [277].

Definition 3.1 (Observationally equivalent parameters [246])**.**

We say that two parameters $\theta_1, \theta_2 \in \Theta$ are *observationally equivalent*, if they encode the same function $f(\cdot;\theta_1) = f(\cdot;\theta_2)$. We denote two observationally equivalent parameters with $\theta_1 \sim \theta_2$.

In the case of homogeneous activations (Assumption 3.1), we describe two kinds of transformations that map a parameter θ into a different but observationally equivalent one.

Neuron rescaling. The input weights of a hidden neuron can be rescaled by a positive scalar $\alpha > 0$ provided that its output weights are rescaled by the inverse α^{-1} (top panel of Figure 3.1a). We formalize this as the action of the multiplicative group $\mathbb{R}_+$ of positive real numbers on Θ_k:

$$\begin{aligned} T\colon \mathbb{R}_+ \times \Theta_k &\to \Theta_k \\ (\alpha, \theta_k) &\mapsto T_\alpha(\theta_k) = \left(\alpha \cdot W_k^{(1)}, \frac{1}{\alpha} \cdot W_k^{(2)}\right). \end{aligned} \tag{3.2}$$

This action can be naturally extended to the space of all parameters by considering the possibility of rescaling all hidden neurons simultaneously by different factors. If $\alpha = (\alpha_1, \ldots, \alpha_l) \in \mathbb{R}_+^l$

$$T_\alpha(\theta) = (\operatorname{diag}(\alpha) W^{(1)}, W^{(2)} \operatorname{diag}(\alpha)^{-1}). \tag{3.3}$$

Proposition 3.2 (Rescaling invariance)**.** *The rescaling group action of Equation* (3.3) *maps a parameter θ to an observationally equivalent one $T_\alpha(\theta) \sim \theta$ for every $\alpha \in \mathbb{R}_+^l$.*

Proof. Let $x \in \mathbb{R}^d$ be an input and $\alpha \in \mathbb{R}^l_+$, we have that

$$\begin{aligned} f(x; T_\alpha(\theta)) &= W^{(2)} \operatorname{diag}(\alpha)^{-1} \sigma\left(\operatorname{diag}(\alpha) W^{(1)} x\right) \\ &= W^{(2)} \operatorname{diag}(\alpha)^{-1} \operatorname{diag}(\alpha) \sigma(W^{(1)} x) \\ &= W^{(2)} \sigma(W^{(1)} x) = f(x; \theta), \end{aligned}$$

because σ is applied element wise and it is positively homogeneous (Assumption 3.1). □

We write $T(\theta)$ to denote the orbit of a parameter θ under the action of T, i.e. the set of all parameters obtained from θ by arbitrarily rescaling its neurons $T(\theta) = \left\{T_\alpha(\theta) |\ \alpha \in \mathbb{R}^l_+\right\}$.

Permutations of the neurons. Besides rescaling, we can obtain an observationally equivalent network by permuting the hidden neurons in a way that preserves their input and output weights (bottom panel of Figure 3.1a).

Given the symmetric group on l elements $\mathcal{S}_l$ of the permutations of $\{1, \dots, l\}$, we write the action

$$\begin{aligned} P \colon \mathcal{S}_l \times \Theta &\to \Theta \\ (\pi, \theta) &\mapsto P_\pi(\theta) = (R_\pi W^{(1)}, W^{(2)} R_\pi^\top). \end{aligned} \tag{3.4}$$

where R_π is the $l \times l$ row-permutation matrix associated to the permutation π.

Proposition 3.3. *The permutation group action of Equation (3.4) maps a parameter θ to an observationally equivalent one $P_\pi(\theta) \sim \theta$ for every $\pi \in \mathcal{S}_l$.*

Proof. Given that the activation function σ is applied component-wise, we have that it commutes with R_π, namely $\sigma(R_\pi x) = R_\pi \sigma(x)$. Thus

$$\begin{aligned} f(x; P_\pi(\theta)) &= W^{(2)} R_\pi^\top \sigma(R_\pi W^{(1)} x) \\ &= W^{(2)} R_\pi^\top R_\pi \sigma(W^{(1)} x) \\ &= W^{(2)} \sigma(W^{(1)} x) = f(x; \theta), \end{aligned}$$

because R_π is an orthogonal matrix. □

Having defined these two actions (Equation (3.3) and Equation (3.4)), we say that two parameters θ and θ' are **observationally equivalent by rescalings and permutations** if θ' can be obtained from θ by a finite sequence of actions of T and P. In this case we write $\theta \overset{\text{rp}}{\sim} \theta'$.

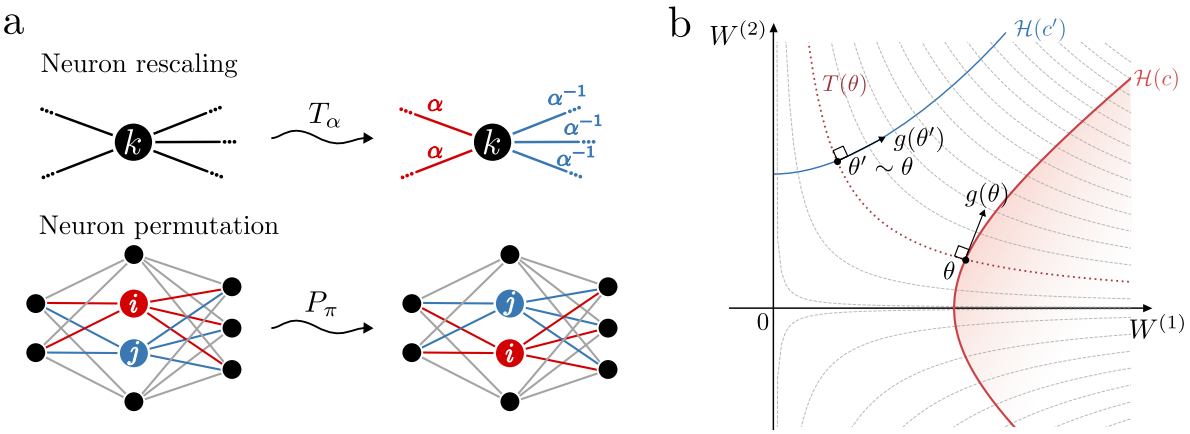

Fig. 3.1 **a.** Depiction of the two group actions acting on the space of the network's parameters: the neuron rescaling of Equation (3.2) (top) and the neuron permutation of Equation (3.4) (bottom). **b.** Depiction of the geometry of the parameter space induced by the rescaling invariance of ReLU networks. The dotted lines denote the orbits $T(\theta)$ while the solid lines represent the invariant sets $\mathcal{H}(c)$ associated with θ and the one associated with its rescaled version θ'. Notice how the gradient of the loss $g(\theta)$ is tangent to $\mathcal{H}(c)$ and orthogonal to $T(\theta)$.

3.2.3 Conserved quantities and the invariant hyperquadrics

The presence of symmetries in the neural network's parameter-function map gives a specific geometric structure to the loss landscape. Let $D = \{(x_i, y_i) \in \mathbb{R}^d \times \mathbb{R}^e\}_{i=1}^N$ be a training dataset of N input-output pairs and fix a loss function $L : \Theta \to \mathbb{R}$ which depends on the parameters only through the output of the neural network (3.1), that is

$$L(\theta) = \frac{1}{N} \sum_{i=1}^{N} \ell(f(x_i; \theta), y_i) \tag{3.5}$$

where $\ell : \mathbb{R}^e \times \mathbb{R}^e \to \mathbb{R}$ is differentiable. This includes, for instance, the standard *mean square error* loss (MSE) with $\ell(x, y) = \|x - y\|^2$.

In this chapter, we consider as a learning process the continuous time version of the gradient descent (GD) algorithm (with learning rate $h > 0$)

$$\theta_{t+1} = \theta_t - h\nabla_\theta L(\theta_t) \tag{3.6}$$

which is named **gradient flow** (GF), and is defined as

$$\frac{\mathrm{d}}{\mathrm{d}t}\theta(t) \in -\nabla_\theta L(\theta(t)) := -g(\theta(t)) \tag{3.7}$$

where $\nabla_\theta L(\theta(t))$ is the Clarke sub-differential [281] which takes into account the parameters θ where $L(\theta)$ is non-differentiable.

The fundamental observation is that the loss function L depends on the parameters only through f and thus its value at θ must be constant over the orbit $T(\theta)$. We can apply this to derive the following important parameter-gradient balance equation.

Proposition 3.4 (Balance equation [253, 250])**.** *For any $\theta = (W^{(1)}, W^{(2)})$ such that L is differentiable at θ, let $g(\theta) = g = (g^{(1)}, g^{(2)})$ be the gradient of the loss function L at θ. Then, it holds that*

$$\sum_{i=1}^{d} W_{ki}^{(1)} g_{ki}^{(1)} - \sum_{j=1}^{e} W_{jk}^{(2)} g_{jk}^{(2)} = 0 \ \ \forall k = 1, \dots, l. \tag{3.8}$$

Proof. In order to prove this result, we assume that θ is a differentiable point for L. If it is not differentiable, the result may still be proved in a different way, as explained in [252].

Let us consider the rescaling action on a single neuron k (Equation (3.2)) $T_{\alpha;k}(\theta) = (\mathrm{diag}(1, \dots, 1, \underbrace{\alpha}_{k}, 1, , \dots, 1)W^{(1)}, W^{(2)} \mathrm{diag}(1, \dots, 1, \underbrace{\alpha^{-1}}_{k}, 1, , \dots, 1))$.

Being a rescaling, we know from Proposition 3.2 that $L(T_{\alpha;k}(\theta)) = L(\theta)\ \forall \alpha \in \mathbb{R}_+$. We now take derivative w.r.t. α at $\alpha = 1$ and derive the following by applying the chain rule.

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\bigg|_{\alpha=1} L(T_{\alpha;k}(\theta)) = 0 \implies \left\langle \nabla L(T_{\alpha;k}(\theta)), \frac{\mathrm{d}}{\mathrm{d}\alpha} T_{\alpha;k}(\theta) \right\rangle \bigg|_{\alpha=1} = 0 \tag{3.9}$$

$$\implies \left\langle g(\theta), \frac{\mathrm{d}}{\mathrm{d}\alpha}\bigg|_{\alpha=1} T_{\alpha;k}(\theta) \right\rangle = 0 \tag{3.10}$$

Computing the second term, we have that

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\alpha}\bigg|_{\alpha=1} T_{\alpha;k}(\theta) &= \frac{\mathrm{d}}{\mathrm{d}\alpha}\bigg|_{\alpha=1} (\mathrm{diag}(1, \dots, 1, \alpha, 1, \dots, 1)W^{(1)}, W^{(2)} \mathrm{diag}(1, \dots, 1, \frac{1}{\alpha}, 1, \dots, 1)) \\
&= (e_{kk} W^{(1)}, W^{(2)} \mathrm{diag}(0, \dots, 0, -\frac{1}{\alpha^2}, 0, \dots, 0))\Big|_{\alpha=1} \\
&= (e_{kk} W^{(1)}, -W^{(2)} e_{kk}) = (W_k^{(1)}, -W_k^{(2)}).
\end{aligned}$$

Plugging this result into Equation (3.10) we finally find that

$$\left\langle g(\theta), \frac{\mathrm{d}}{\mathrm{d}\alpha}\Big|_{\alpha=1} T_{\alpha;k}(\theta) \right\rangle = 0 \iff \left\langle (g^{(1)}, g^{(2)}), (W_k^{(1)}, -W_k^{(2)}) \right\rangle = 0$$
$$\iff \left\langle g^{(1)}, W_k^{(1)} \right\rangle - \left\langle g^{(2)}, W_k^{(2)} \right\rangle = 0$$

that, when the Frobenius inner product is expanded, results in Equation (3.8). □

Another way to get to this result is to start by the fact that the gradient of a function at a point is always orthogonal to the level set at that point. Given that orbits under rescaling are contained in the level set, we find that the gradient must be orthogonal to the orbits (as pictured in Figure 3.1b).

For convenience of notation, we define, for any $k = 1, \dots, l$, the following bilinear forms on Θ, which help us describe compactly the geometry induced by the rescaling symmetry.

Definition 3.2 (Balance form)**.**

If $\theta = (W^{(1)}, W^{(2)})$ and $\eta = (U^{(1)}, U^{(2)})$, we define the **balance form** as the bilinear form

$$\langle\langle \theta, \eta \rangle\rangle_k := \sum_{i=1}^{d} W_{ki}^{(1)} U_{ki}^{(1)} - \sum_{j=1}^{e} W_{jk}^{(2)} U_{jk}^{(2)} \tag{3.11}$$

which, notice, only depends on the k-th row of $W^{(1)}$ and k-th column of $W^{(2)}$ meaning that we can equivalently see it as a bilinear form on Θ_k. We call the value $\langle\langle \theta, \theta \rangle\rangle_k$ the **balance** of parameter θ at hidden neuron k.

With the notation given by Definition 3.2, we see that Proposition 3.4 can be simply rewritten as $\langle\langle \theta, g(\theta) \rangle\rangle_k = 0$ for every neuron k. These conditions put constraints on the possible values of the gradient and, by extension, on the possible gradient flow trajectories. We write them in terms of *conservation laws* under the learning dynamics.

Proposition 3.5 (Gradient flow conservation law [252, 251])**.** *Under gradient flow optimization (Equation (3.7)), if L is differentiable at θ, we have that*

$$\frac{\mathrm{d}}{\mathrm{d}t} \langle\langle \theta, \theta \rangle\rangle_k = 0 \; \forall k = 1, \dots, l. \tag{3.12}$$

Proof. Under gradient flow optimization $\dot{\theta} = -g(\theta)$, therefore

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle\langle\theta,\theta\rangle\rangle_k = 2\langle\langle\dot{\theta},\theta\rangle\rangle_k = -2\langle\langle g(\theta),\theta\rangle\rangle_k = 0 \;\; \forall k = 1,\dots,l. \tag{3.13}$$

□

This result, first obtained in [252] and discussed in [253, 251], tells us that the rescaling symmetry results in the quantities $\langle\langle\theta,\theta\rangle\rangle_k$ being conserved. This means that the difference between the Euclidean norm of the inputs and the outputs weights to every neuron is constant throughout the GF training trajectory. Moreover, under the condition of homogeneity of the activation function, [252] proves that Equation (3.12) holds even at non-differentiable points of L. In Chapter 4 we generalize these conservation laws to any feed-forward neural network architecture.

Invariant sets. Assume that, at the initialization θ_0, we have $\langle\langle\theta_0,\theta_0\rangle\rangle_k = c_k$, for all k, then Equation (3.12) implies that the GF trajectory will lie on the set characterized by the system of equations $\langle\langle\theta,\theta\rangle\rangle_k = c_k$ for $k = 1,\dots,l$. This subset of the parameter space Θ is mapped to itself under the GF dynamics by Equation (3.12) (see Figure 3.1b) and constitutes the main object of our study.

Definition 3.3 (Invariant set [264])**.**

Given $c = (c_1,\dots,c_l)$, we call **invariant set** the subset $\mathcal{H}(c) \subseteq \Theta$ given by the equations $\langle\langle\theta,\theta\rangle\rangle_k = c_k \;\forall k = 1,\dots,l$.

If we look at each single equation (i.e. to each hidden neuron), we see that any single equation of the invariant set can be written as

$$\sum_{i=1}^{d}\left(W_{ki}^{(1)}\right)^2 - \sum_{j=1}^{e}\left(W_{jk}^{(2)}\right)^2 = c_k \tag{3.14}$$

which corresponds to a *hyperquadric* (or quadric hypersurface) in Θ_k. We denote with $\mathcal{Q}(c_k) \subseteq \Theta_k$ this hypersurface and call it the **invariant hyperquadric** associated to the k-th hidden neuron.

Here $c_k \in \mathbb{R}$ takes the role of a label associated with the k-th hidden neuron, which, we see in the next section, plays a key role in specifying the shape of $\mathcal{Q}(c_k)$. Figure 3.2a shows how,

for $d = 2$ and $e = 1$, $\mathcal{Q}(c_k)$ is an hyperboloid with one sheet (connected) if $c_k > 0$, two sheets if $c_k < 0$ and it is a cone when $c_k = 0$.

3.3 Topology of the invariant set

As we discussed above, Equation (3.12) tells us that gradient flow trajectories can't explore the whole space Θ but are constrained to lie on the invariant set $\mathcal{H}(c)$. The values of c, in turn, depend on the initialization and, we see from Equation (3.14), quantify the balance between the norms of input and output weights in every hidden neuron.

The goal of this section is to provide a topological characterization of $\mathcal{H}(c)$ that can tell us something about the presence or absence of fundamental *obstructions* to the network's training process. With obstruction, we mean the impossibility of a GF trajectory to travel freely from one point to another in $\mathcal{H}(c)$.

Our topological characterization will be framed using *Betti numbers*. Betti numbers, we recall from Chapter 0, are well-known topological invariants (see Definition 0.7) given by a sequence of natural numbers that intuitively encode the number of higher-dimensional holes and cavities present in space. In particular, the 0-th Betti number of a topological space X, $\beta_0(X)$ corresponds to the number of connected components of X and thus will be fundamental for our goal of identifying obstructions.

The invariant set $\mathcal{H}(c)$ is given as the set of solutions of l polynomial equations of degree 2 sharing no variables. In the setting of two-layer neural networks, we can leverage the fact that the parameter space can be decomposed into the parameter spaces of the hidden neurons. This, in turn, allows us to decompose the invariant set as the product of the neurons' invariant hyperquadrics, greatly simplifying our study.

Lemma 3.1 (Decomposition of the invariant set)**.** *In a two-layer ReLU neural network, the invariant set $\mathcal{H}(c)$ is homeomorphic to the Cartesian product of the hidden neurons' invariant hyperquadrics; that is*

$$\mathcal{H}(c) \cong \mathcal{Q}(c_1) \times \cdots \times \mathcal{Q}(c_l). \tag{3.15}$$

Proof. Proposition 3.1 tells us that the invariant set can be decomposed as the direct sum of the single hidden neurons' parameter spaces. This means that, for every $\theta \in \Theta$, there exist unique $\theta_1 \in \Theta_1, \ldots, \theta_l \in \Theta_l$ such that

$$\theta = \theta_1 + \theta_2 + \cdots + \theta_l.$$

Therefore, we have a linear isomorphism $\varphi : \Theta_1 \times \cdots \times \Theta_k \to \Theta$

$$\varphi : (\theta_1, \ldots, \theta_l) \mapsto \theta_1 + \cdots + \theta_l = \theta.$$

The invariant set is a subset of Θ, which is given as the set of solutions of l equations $\langle\langle \theta, \theta \rangle\rangle_k = c_k \; k = 1, \ldots, l$. Notice that each of these equations involves a set of variables that appear only in that particular equation. These variables are exactly the ones which belong to Θ_k. In fact $\langle\langle \theta, \theta \rangle\rangle_k = \langle\langle \theta_k, \theta_k \rangle\rangle_k$.

Therefore, given $\theta_1 \in Q(c_1), \ldots, \theta_l \in Q(c_l)$ we have that

$$\varphi(\theta_1, \ldots, \theta_l) = \sum_{k=1}^{l} \theta_k \in \mathcal{H}(c).$$

On the opposite, given $\theta \in \mathcal{H}(c)$ we have that

$$\varphi^{-1}(\theta) = (\theta_1, \ldots, \theta_l) \in Q(c_1) \times \cdots \times Q(c_l).$$

Therefore, $\mathcal{H}(c)$ is in bijection with $Q(c_1) \times \cdots \times Q(c_l)$ through φ which, being a linear isomorphism, is also a homeomorphism, implying also that $\mathcal{H}(c)$ and $Q(c_1) \times \cdots \times Q(c_l)$ are homeomorphic. □

Lemma 3.1 tells us that we can understand the topology of $\mathcal{H}(c)$ by studying its factors independently. These, the hyperquadrics, are well-studied objects for which the next proposition provides a full topological characterization.

Proposition 3.6 (Topology of the invariant hyperquadrics)**.** *If $c_k > 0$, $Q(c_k)$ is a topological manifold homeomorphic to $\mathbb{R}^e \times S^{d-1}$. If $c_k < 0$, $Q(c_k)$ is a topological manifold homeomorphic to $\mathbb{R}^d \times S^{e-1}$. If $c_k = 0$, $Q(0)$ is a contractible space.*

Proof. Let us consider the three cases separately. In the first two cases we will explicitly find the homeomorphism, while in the third we will exhibit a deformation retraction (Definition 0.11) from $Q(0)$ to a point.

- If $c_k > 0$, $Q(c_k)$ is defined by the equation

$$\sum_{i=1}^{d} \left(W_{ki}^{(1)}\right)^2 - \sum_{j=1}^{e} \left(W_{jk}^{(2)}\right)^2 = c_k \iff \left\|W_k^{(1)}\right\|_F^2 - \left\|W_k^{(2)}\right\|_F^2 = c_k,$$

where $\|\cdot\|_F$ is the Frobenius norm of a matrix, i.e., the square root of the sum of the squares of its elements. This can be rewritten as

$$\left\|W_k^{(1)}\right\|_F = \sqrt{c_k + \left\|W_k^{(2)}\right\|_F^2} \tag{3.16}$$

where $c_k + \left\|W_k^{(2)}\right\|_F^2 > 0$ because $c_k > 0$. We define the map $h : Q(c_k) \to S^{d-1} \times \mathbb{R}^e$ as

$$h\left(W_k^{(1)}, W_k^{(2)}\right) = \left(\frac{W_k^{(1)}}{\sqrt{c_k + \left\|W_k^{(2)}\right\|_F^2}}, W_k^{(2)}\right) \tag{3.17}$$

where, notice, the first component belongs to the sphere S^{d-1} because of Equation (3.16) and $W_k^{(2)} \in \mathbb{R}^e$. This map is bijective, differentiable and has the following inverse $h^{-1} : S^{d-1} \times \mathbb{R}^e \to Q(k)$

$$h^{-1}(u, x) = (\sqrt{c_k + \|x\|_F^2}\, u, x)$$

which is differentiable. Therefore, h is a diffeomorphism from $Q(c_k)$ to $S^{d-1} \times \mathbb{R}^e$.

• If $c_k < 0$, we write the equation of $Q(c_k)$ as

$$\left\|W_k^{(2)}\right\|_2 = \sqrt{-c_k + \left\|W_k^{(1)}\right\|_2^2}$$

where $W_k^{(1)}$ and $W_k^{(2)}$ have switched their roles to guarantee the term under the square root to be positive. The diffeomorphism is now built analogously to Equation (3.17) as a map $h : Q(c_k) \to S^{e-1} \times \mathbb{R}^d$.

• If $c_k = 0$, we prove that $Q(0)$ is a contractible space. To do that, we show that there exists a deformation retraction of $Q(0)$ onto the point $0 \in Q(0)$, i.e. a continuous map $h : [0, 1] \times Q(0) \to Q(0)$ such that $h(0, \theta_k) = \theta_k$ and $h(1, \theta_k) = 0\ \forall \theta_k \in Q(0)$. This map performs a linear scaling of $Q(0) = \{\theta_k \in \Theta_k | \langle\langle \theta_k, \theta_k \rangle\rangle_k = 0\}$ to the origin and is defined as

$$h(\lambda, \theta_k) = (1 - \lambda)\theta_k.$$

This is continuous and well-defined because

$$\langle\langle h(\lambda, \theta_k), h(\lambda, \theta_k) \rangle\rangle_k = \langle\langle (1 - \lambda)\theta_k, (1 - \lambda)\theta_k \rangle\rangle_k = (1 - \lambda)^2 \langle\langle \theta_k, \theta_k \rangle\rangle_k = 0,$$

meaning that $h(\lambda, \theta_k) \in Q(0)$ for every $\theta_k \in Q(0)$ and for every $\lambda \in [0, 1]$. □

Leveraging the decomposition of Lemma 3.1 and the characterization of the factors given by Proposition 3.6, we can now give a complete topological characterization of the invariant set of a two layer ReLU neural network.

Theorem 3.1 (Invariant set topology)**.** *Let θ_0 be the initialization of the network and $c \in \mathbb{R}^l$, with $c_k = \langle\langle \theta_0, \theta_0 \rangle\rangle_k$. Let l_+, l_-, l_0 be the number of positive, negative, and zero components of c. Then, the invariant set is homeomorphic to the following Cartesian product:*

$$\mathcal{H}(c) \cong (\mathbb{R}^e \times S^{d-1})^{\times l_+} \times (\mathbb{R}^d \times S^{e-1})^{\times l_-} \times Q(0)^{\times l_0}.$$

Proof. The result follows by combining the decomposition of $\mathcal{H}(c)$ into the invariant hyperquadrics (Lemma 3.1) and their topological characterizations (Proposition 3.6). □

It follows from Theorem 3.1 that we can explicitly compute all the Betti numbers (Definition 0.7) of the invariant set. We give the following result in terms of the *Poincaré polynomial* of $\mathcal{H}(c)$, namely the polynomial whose coefficients are the Betti numbers (see Definition 0.9).

Corollary 3.1 (Poincaré polynomial of the invariant set)**.** *Let θ_0 be the initialization of the network and $c \in \mathbb{R}^l$, with $c_k = \langle\langle \theta_0, \theta_0 \rangle\rangle_k$. Let l_+, l_-, l_0 be the number of positive, negative, and zero components of c, respectively. The Poincaré polynomial of $\mathcal{H}(c)$ is given by*

$$p_{\mathcal{H}(c)}(t) = (1 + t^{d-1})^{l_+}(1 + t^{e-1})^{l_-} \tag{3.18}$$

Proof. To prove the result, we leverage the corollary of the Künneth formula (Corollary 0.1) that tells us that the Poincaré polynomial of the Cartesian product of two spaces is equal to the product of their Poincaré polynomials.

Starting from Proposition 3.6, we can apply this result to $Q(c_k)$.

$$p_{Q(c_k)}(t) = \begin{cases} p_{\mathbb{R}^e}(t)\, p_{S^{d-1}}(t) & \text{if } c_k > 0 \\ p_{\mathbb{R}^d}(t)\, p_{S^{e-1}}(t) & \text{if } c_k < 0 \\ 1 & \text{if } c_k = 0 \end{cases} \tag{3.19}$$

because a contractible space has 1 connected component and all of its other Betti numbers equal to zero. Moreover, we know that $\mathbb{R}^n$ is contractible for any n and its Poincaré polynomial is therefore $p_{\mathbb{R}^n}(t) = 1$. The Poincaré polynomial of the sphere S^n is given by $p_{S^n}(t) = 1 + t^n$.

Equation (3.19) becomes

$$p_{\mathcal{Q}(c_k)}(t) = \begin{cases} 1 + t^{d-1} & \text{if } c_k > 0 \\ 1 + t^{e-1} & \text{if } c_k < 0 \\ 1 & \text{if } c_k = 0 \end{cases}. \tag{3.20}$$

Given that Lemma 3.1 tells us that $\mathcal{H}(c)$ can be factored into the product of the $\mathcal{Q}(c_k)$, we apply Künneth formula and find that

$$p_{\mathcal{H}(c)}(t) = p_{\mathcal{Q}(c_1)}(t) \cdots p_{\mathcal{Q}(c_l)}(t) = (1 + t^{d-1})^{l_+}(1 + t^{e-1})^{l_-}. \tag{3.21}$$

□

Corollary 3.1 contains a wealth of topological information, as it gives us the exact number of holes and cavities of any order, depending on the network's structural hyperparameters (d, e) and initialization (l_+, l_-). In the rest of this chapter, we focus only on the 0-th Betti number, as the non-connectedness of $\mathcal{H}(c)$ provides a clear obstruction to the gradient flow trajectories.

Connectedness of the invariant set. With regard to the connectedness of $\mathcal{H}(c)$, we can leverage Corollary 3.1 to obtain the exact number of connected components.

Corollary 3.2 (Connectedness of the invariant set)**.** *The 0-th Betti number β_0 of $\mathcal{H}(c)$, corresponding to the number of its connected components, is given by*

$$\beta_0 = \begin{cases} 1 & \text{if } d, e > 1 \\ 2^{l_+} & \text{if } d = 1, e > 1 \\ 2^{l_-} & \text{if } d > 1, e = 1 \\ 2^{l_+ + l_-} & \text{if } d = 1, e = 1 \end{cases} \tag{3.22}$$

Proof. This can be directly obtained by computing the coefficient of degree 0 of the Poincaré polynomial obtained through Corollary 3.1. □

What we see in Equation (3.22) is that, in most cases, the invariant set is connected, and gradient flow has no inherent topological limitations in exploring the whole of $\mathcal{H}(c)$. Instead, when the hidden neurons have only one input or only one output, the space is fragmented into several components whose number scales exponentially in l_+ or l_-, respectively.

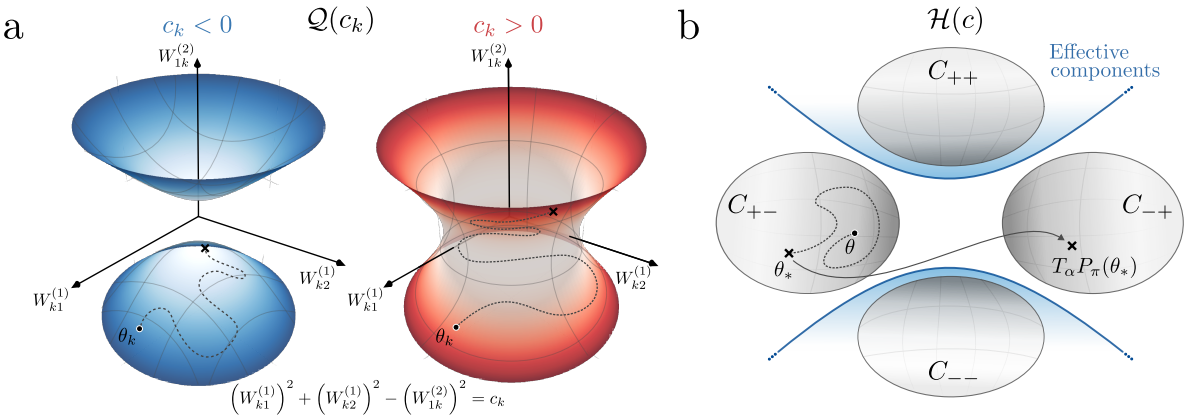

Fig. 3.2 **a.** The invariant hyperquadric $\mathcal{Q}(c_k)$ of a neuron with two inputs ($d = 2$) and one output ($e = 1$) in the cases where $c_k < 0$ (left) and $c_k > 0$ (right). **b.** Depiction of the invariant set $\mathcal{H}(c)$ in the case where $l_- = 2$ so that there are $2^{l_-} = 4$ connected components. $C_{\pm\mp}$ denotes the connected component such that $s = (\pm 1, \mp 1)$. The blue lines separate the different effective components of $\mathcal{H}(c)$.

Let us focus on the more interesting case where $d > 1$ and $e = 1$, i.e. there is a single output neuron. This happens in standard machine learning scenarios like scalar regression and binary classification, where one tries to learn a function with a single scalar output.

Corollary 3.3. *If the output of a two-layer ReLU neural network is a single scalar $e = 1$, its input has dimension $d > 1$, and the initial parameter θ_0 is such that $\langle\langle \theta_0, \theta_0 \rangle\rangle_k < 0$ for $l_- > 0$ hidden neurons, then the set $\mathcal{H}(c)$ is disconnected and has 2^{l_-} connected components.*

This means that neurons initialized with the norm of their outgoing weight strictly greater than their incoming weights' norm are responsible for disconnecting the space, and we now precisely identify which connected component a parameter θ belongs to and clarify the meaning of the obstruction.

Proposition 3.7 (Connected components characterization)**.** *Let $e = 1, d > 1$, and $\theta \in \mathcal{H}(c)$ with c such that $c_{k_1}, \dots, c_{k_{l_-}} < 0$ while $c_k \geq 0$ for all other k. Let $W_-^{(2)} := (W_{k_1}^{(2)}, \dots, W_{k_{l_-}}^{(2)}) \in \mathbb{R}^{1 \times l_-}$ be the row vector whose components are the components of $W^{(2)} \in \mathbb{R}^{1 \times l}$ associated to $c_k < 0$. Then the vector $s(\theta) = (\operatorname{sign}(W_{k_1}^{(2)}), \dots, \operatorname{sign}(W_{k_{l_-}}^{(2)}))$ identifies uniquely the component θ belongs to, namely: θ and θ' belong to the same connected component of $\mathcal{H}(c)$ if and only if $s(\theta) = s(\theta')$.*

Proposition 3.7, which we prove below, implies that the vector $s(\theta)$ containing the signs of the output weights to negatively balanced neurons does not change when we move in a connected component on a continuous curve such as the one given by gradient flow. This gives

us an interesting interpretation of the topological obstruction: *gradient flow cannot change the signs of the outgoing weights of the hidden neurons k such that $c_k < 0$.* Proposition 3.7 extends one of the results of Ref. [259] which proves that the same also holds when $c_k = 0$ (balanced initialization).

By also considering Corollary 3.2, one obtains that a clever initialization of the parameters given by $\langle\langle \theta_0, \theta_0 \rangle\rangle_k \geq 0 \; \forall k = 1, \dots, l$ can prevent the issue by ensuring the connectedness of the invariant set.

Proof of Proposition 3.7. In the following, we exploit Lemma 3.2 and use the terminology connected and path-connected interchangeably.

• Let us first prove that $s(\theta) = s(\theta')$ means that θ and θ' belong to the same connected component. We do this by explicitly building a continuous curve $\delta : [0, 1] \to \mathcal{H}(c)$ such that $\delta(0) = \theta$ and $\delta(1) = \theta'$. Let us proceed by leveraging the homeomorphism from $\mathcal{H}(c)$ to $\mathcal{Q}(c_1) \times \cdots \times \mathcal{Q}(c_l)$ (Lemma 3.1) and consider the different components of δ in the invariant hyperquadrics associated to each neuron $\delta = (\delta_1, \dots, \delta_l)$.

If $c_k \geq 0$, we know that $\mathcal{Q}(c_k)$ is path-connected and therefore we take $\delta_k(t)$ to be any continuous curve in $\mathcal{Q}(c_k)$ such that $\delta_k(0) = \theta_k$ and $\delta_k(1) = \theta'_k$.

When $c_k < 0$, we know that the invariant hyperquadric $\mathcal{Q}(c_k)$ has two connected components. We build a continuous curve joining θ_k to θ'_k by defining a curve in $\mathcal{Q}(c_k)$ that maps any point θ_k to one of the two vertices of $\mathcal{Q}(c_k)$. If $c_k < 0$ for $k \in= \{k_1, \dots, k_{l_-}\} \subseteq \{1, \dots, l\}$, we define the curve $\gamma_{i;\theta} : [0, 1] \to \mathcal{Q}(c_{k_i})$ with

$$\gamma_{i;\theta}(t) = \left((1-t) W^{(1)}_{k_i 1}, \dots, (1-t) W^{(1)}_{k_i l}, s(\theta)_i \sqrt{-c_{k_i} + (1-t)^2 \left\| W^{(1)}_{k_i} \right\|^2_F} \right)$$

for every $i = 1, \dots, l_-$.

$\gamma_{i;\theta}$ is a continuous curve which belongs to $\mathcal{Q}(c_k)$ and connects the point $\gamma_{i;\theta}(0) = \theta_{k_i}$ with $\gamma_{i;\theta}(1) = (0, s(\theta)_i \sqrt{-c_{k_i}})$, i.e. the hyperquadric's vertex.

If we define $\bar{\gamma}_{i;\theta}(t) := \gamma_{i;\theta}(1-t)$, which is the same curve as $\gamma_{i;\theta}$ but traversed in the opposite direction, we can define the concatenated curve

$$\delta_{k_i}(t) = \bar{\gamma}_{i;\theta'} \gamma_{i;\theta}(t)$$

i.e. the curve which travels on $\gamma_{i;\theta}$ for $t \in [0, \frac{1}{2}]$ and on $\bar{\gamma}_{i;\theta'}$ for $t \in [\frac{1}{2}, 1]$, for $i = 1, \dots, l_-$. Notice now that $\delta_{k_i}(0) = \theta_{k_i}$ and $\delta_{k_i}(1) = \theta'_{k_i}$. Moreover δ_{k_i} is continuous because

$$\gamma_{i;\theta}(1) = (0, s(\theta)_i\sqrt{-c_{k_i}}) = (0, s(\theta')_i\sqrt{-c_{k_i}}) = \bar{\gamma}_{i;\theta'}(0)$$

under the hypothesis that $s(\theta) = s(\theta')$.

Finally, we found a continuous curve $\delta = (\delta_1, \dots, \delta_l)$ such that $\delta(t) \in \mathcal{H}(c)\ \forall t \in [0, 1]$ and $\delta(0) = \theta,\ \delta(1) = \theta'$. Therefore, θ and θ' belong to the same connected (path) component.

• Let us now prove that if θ and θ' belong to the same connected component, then $s(\theta) = s(\theta')$.

If θ, θ' belong to the same path component, there exists a continuous curve connecting them. Let $\gamma : [0, 1] \to \mathcal{H}(c)$ be such a continuous curve in $\mathcal{H}(c)$ such that $\gamma(0) = \theta$ and $\gamma(1) = \theta'$.

For each $k_i \in \{k_1, \dots, k_{l_-}\}$ such that $c_{k_i} < 0$, we know that $\gamma_{k_i}(t) \in \mathcal{Q}(c_k)$ means that

$$\gamma_{k_i}(t) = \left(\gamma^{(1)}_{k_i 1}(t), \dots, \gamma^{(1)}_{k_i d}(t), \underbrace{\sigma_{k_i}(t)\sqrt{-c_{k_i} + \left\|\gamma^{(1)}_{k_i}(t)\right\|^2_F}}_{\gamma^{(2)}_{k_i}(t)}\right)$$

for some function $\sigma_{k_i}(t) \in \{-1, 1\}$ such that $\sigma_{k_i}(0) = s(\theta)_i$ and $\sigma_{k_i}(1) = s(\theta')_i$.

Assume, by contradiction, that $s(\theta)_i = -s(\theta')_i$. Assume also, without loss of generality, that $s(\theta)_i = +1$ and $s(\theta')_i = -1$. The function

$$\gamma^{(2)}_{k_i}(t) = \sigma_{k_i}(t)\sqrt{-c_{k_i} + \left\|\gamma^{(1)}_{k_i}(t)\right\|^2_F},$$

then, is a continuous function such that $\gamma^{(2)}_{k_i}(0) > 0$ and $\gamma^{(2)}_{k_i}(1) < 0$ and thus, by the intermediate value theorem, there exists $t_* \in (0, 1)$ such that $\gamma^{(2)}_{k_i}(t_*) = 0$.

But $\gamma^{(2)}_{k_i}(t) \neq 0$ for every t, as $\sigma_{k_i}(t) \in \{-1, 1\}$ and $\sqrt{-c_{k_i} + \left\|\gamma^{(1)}_{k_i}\right\|^2_F} > 0$, leading to a contradiction. Repeating the argument, we then prove by contradiction that $s(\theta)_{k_i} = s(\theta')_{k_i}$ for all $k \in k_1, \dots, k_{l_-}$, thus concluding the proof. □

3.4 Taking permutation symmetries into account

Corollary 3.3 states that neurons k such that $c_k < 0$ are "pathological", in the sense that they are responsible for disconnecting the invariant set into several components whose number scale exponentially in the number of those neurons. This result gives us a grim picture of the possibility of actually optimizing the neural network: if the initial parameter θ_0 is in a particular connected component and the global optimum θ_* lies in another, then any gradient flow trajectory will not be able to reach θ_* because it will be constrained in its connected component.

This result, however, provides us only with a partial picture of the parameter space's geometry. It is a priori possible that the training trajectory, moving in its connected component, reaches a parameter $\hat{\theta}$, which itself is optimal as it is observationally equivalent (Definition 3.1) to θ_* ($\hat{\theta} \sim \theta_*$). In this case, the topological obstruction given by the non-connectedness would be only apparent.

To take this fact into account, we define the following notion.

Definition 3.4 (Effective component)**.**

Let $\theta \in \mathcal{H}(c)$ and $C(\theta)$ be its connected component. We define its effective component $\mathrm{Eff}(\theta)$ as the union of the connected component of all θ' such that $\theta' \overset{\mathrm{rp}}{\sim} \theta$. So that

$$\mathrm{Eff}(\theta) := \bigcup_{\theta' \overset{\mathrm{rp}}{\sim} \theta} C(\theta').$$

Figure 3.2b gives a picture that clarifies the definition: a space with 4 connected components has only 3 effective components. If the optimum θ_* belongs to the same effective component as the initialization, then it is possible to reach a parameter that is observationally equivalent to it (through permutations and rescalings).

Before characterizing the effective components, we present here a useful result that tells us that the rescaling action (Equation (3.3)) can take any non-degenerate parameter $\theta \in \mathcal{H}(c)$ to any other invariant set $\mathcal{H}(c')$ for every $c' \in \mathbb{R}^l$. This means that all the invariant sets are *functionally equivalent* (they contain the same functions) even if they may not be *topologically equivalent*.

Proposition 3.8 (Rescaling equivalence of invariant sets)**.** *For every $c_k \in \mathbb{R}$ and for every $\theta_k \in \Theta_k$ such that $W_k^{(1)}, W_k^{(2)} \neq 0$, there exists a unique $\alpha_k \in \mathbb{R}_+$ such that $T_{\alpha_k}(\theta_k) \in Q(c_k)$. If*

$W_k^{(1)} = 0$ *and* $W_k^{(2)} \neq 0$*, then the same holds for every* $c_k < 0$*, while, if* $W_k^{(1)} \neq 0$ *and* $W_k^{(2)} = 0$*, it holds for every* $c_k > 0$*.*

The proof can be found in Section 3.9.1, together with the formula of the specific α which realizes the rescaling.

The following theorem leverages the power of Proposition 3.8 to give necessary and sufficient conditions for θ and θ' to belong to the same effective component.

Theorem 3.2 (Characterization of effective components)**.** *Let* $d > 1$ *and* $e = 1$*. Let* $c \in \mathbb{R}^l$ *and* l_- *be the number of neurons such that* $c_k < 0$*. Assume that* $l_- \geq 1$*. Let* $C, C' \subseteq \mathcal{H}(c)$ *be two distinct connected components of* $\mathcal{H}(c)$ *such that* $s(\theta) = s$*,* $\forall \theta \in C$*, and* $s(\theta') = s'$*,* $\forall \theta' \in C'$*. Then, the following statements are equivalent:*

1. *for every* $\theta \in C$ *there exists* $\theta' \in C'$ *such that* $\theta \overset{\text{rp}}{\sim} \theta'$*;*
2. $\sum_{i=1}^{l_-} s_i = \sum_{i=1}^{l_-} s'_i$

Proof. • Let us first prove that, if for every $\theta \in C$ there exists a $\theta' \in C'$ such that $\theta \overset{\text{rp}}{\sim} \theta'$ then $\sum_{i=1}^{l_-} s(\theta)_i = \sum_{i=1}^{l_-} s(\theta')_i$.

First, Lemma 3.3 tells us that we can interchange rescaling and permutation if we permute the rescaling factors accordingly. This means we can reduce any composite action of rescalings and permutations to the action of a single rescaling and a single permutation.

Let $\theta \in C$ and $\theta' \in C'$ such that $\theta \overset{\text{rp}}{\sim} \theta'$. Then there exist $\alpha \in \mathbb{R}_+^l$ and $\pi \in \mathcal{S}_l$ such that

$$T_\alpha P_\pi(\theta) = \theta'. \tag{3.23}$$

This means that $T_\alpha P_\pi(\theta)$ and θ' belong to the same invariant set and, specifically, to the same connected component. Therefore,

$$s(T_\alpha P_\pi(\theta)) = s(\theta').$$

Notice that

$$s(T_\alpha P_\pi(\theta)) = s(P_\pi(\theta))$$

because T_α does not change the sign of $W^{(2)}$ as it acts by scaling it by positive factors. Let us focus on

$$s(P_\pi(\theta)) = \text{sign}((W^{(2)} R_\pi^\top)_-).$$

If the neurons of θ such that $c_k < 0$ are indexed by $k_1, k_2, \ldots, k_{l_-}$, we will have that the neurons of $P_\pi(\theta)$ such that $c_k < 0$ are indexed by $\pi(k_1), \pi(k_2), \ldots, \pi(k_{l_-})$. Therefore

$$s(P_\pi(\theta))_i = \operatorname{sign}(W^{(2)}_{\pi(k_i)}) = (s(\theta)R^\top_{\pi_-})_i$$

for some permutation $\pi_- \in \mathcal{S}_{l_-}$. Therefore, Equation (3.23) implies that

$$s(\theta') = s(P_\pi(\theta)) = s(\theta)R^\top_{\pi_-}.$$

The action of rescaling and permutation can only reshuffle the label s of the connected component. This means that

$$s(\theta)R^\top_{\pi_-} = s(\theta') \implies \sum_{i=1}^{l_-}(s(\theta)R^\top_{\pi_-})_i = \sum_{i=1}^{l_-} s(\theta')_i \implies \sum_{i=1}^{l_-} s(\theta)_i = \sum_{i=1}^{l_-} s(\theta')_i.$$

• Let us now prove the other implication. Let $s, s' \in \mathbb{R}^{l_-}$ such that $\sum_{i=1}^{l_-} s_i = \sum_{i=1}^{l_-} s'_i$.

Given that their sum is equal, s and s' have the same number of $+1$ and -1 and thus there exists a permutation $\pi_- \in \mathcal{S}_{l_-}$ such that $s' = sR^\top_{\pi_-}$.

Let $\pi \in \mathcal{S}_l$ be the permutation which permutes the neurons such that $c_k < 0$ according to π_- and leaves the others fixed. In this way $s(P_\pi(\theta)) = sR^\top_{\pi_-} = s'$.

$P_\pi(\theta)$, however, doesn't belong to $\mathcal{H}(c)$ but to another invariant set given by $\mathcal{H}(R_\pi c)$.

Applying Proposition 3.8 we can find a rescaling $\alpha = \alpha(\pi) \in \mathbb{R}^l_+$ such that $T_\alpha P_\pi(\theta) \in \mathcal{H}(c)$.

Since neurons such that $c_k \geq 0$, are left unchanged by the permutation, we can apply the proposition and we rescale them with $\alpha_k = 1$. The permuted neurons are the ones such that $c_k < 0$, namely the ones whose weights satisfy $\left\|W^{(1)}_k\right\|^2_F - \left\|W^{(2)}_k\right\|^2_F < 0$, meaning that $W^{(2)}_k \neq 0$.

As noted above, the action of the rescaling doesn't change the sign vector, and thus

$$s(T_\alpha P_\pi(\theta)) = s(P_\pi(\theta)) = s'.$$

If we name $\theta' := T_\alpha P_\pi(\theta)$ this result means that we found a $\theta' \overset{\text{rp}}{\sim} \theta$ such that $\theta' \in C'$, thus concluding the proof. □

Theorem 3.2 tells us that, while connected components are identified by s, the effective components are identified only by the values of $\sum_i s_i$ or, equivalently, by the number of positve and negative terms in s. Therefore, we find that the number of effective components scales much slower than the exponential growth of the number of connected components given by Corollary 3.2.

Corollary 3.4 (Number of effective components)**.** *The number of effective components of* $\mathcal{H}(c)$ *is given by* $1 + l_-$.

Proof. Theorem 3.2 tells us that two connected components C, C' belong to the same effective component if and only if their associated sign vectors $s, s' \in \{-1, 1\}^{l_-}$ have the same sum. The number of effective components will thus equal the number of different values that the sum $\sum_{i=1}^{l_-} s_i$ can have. If $s_i = 1\ \forall i$ then $\sum_{i=1}^{l_-} s = l_-$. Each switch of a component to -1 decreases the sum's value by 2 until it reaches the minimum $-l_-$. Therefore, the total number of values of the sum will be $1 + l_-$. □

3.5 Empirical Validation

Task, dataset, and model setup. We display here a toy example, showing how the initialization of the model can cause a topological obstruction that makes the optimum unreachable.

We consider the simple linear function $f^*(x_1, x_2) = -(x_1 + x_2)$, which we want our neural network to learn. Next, we generate a dataset of 8000 points $(x_i, f^*(x_i))$ by sampling $x_i \sim U([0, 1]^2)$. Our model, depicted in Figure 3.3a) is a one hidden layer neural network with 2 hidden neurons, ReLU activations and no biases. All the weights are initialized by independently sampling from $U(\left[-\sqrt{2}, \sqrt{2}\right])$. From the task and the network's architecture, it is clear that at least one of the output weights has to be negative to approximate f^* correctly.

To standardize our results, we apply the rescaling of Proposition 3.8 and relocate the initial parameters to an observationally equivalent one in the invariant set $\mathcal{H}(c)$ with $c_k \in \{-0.1, 0.1\}$, controlling the sign of the weights on the last layer. We allow ourselves to do these two manipulations to control the experiments while only marginally modifying the network initialization, avoiding the introduction of massively unbalanced weights, which could change the dynamics, as shown in [282]. Finally, we train the network using gradient descent on the MSE loss with a small learning rate of $h = 0.01$. This limits the variations of c_k values to less than one percent along the training, giving us a good approximation of the gradient flow dynamics (Equation (3.7)).

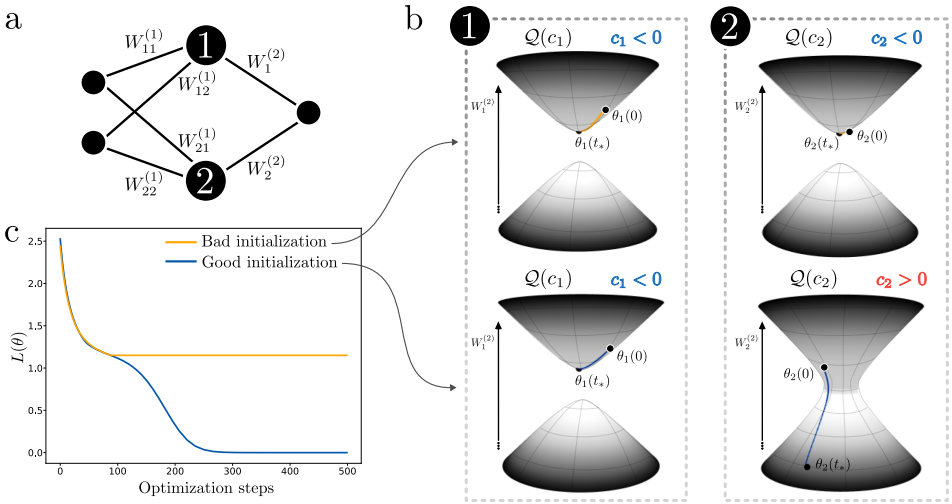

Fig. 3.3 Visualization of the experimental setup described in Section 3.5. **a.** The small 2-layer neural network architecture considered. **b.** The hidden neurons' parameter spaces, together with the invariant hyperquadrics associated with hidden neurons 1 (left) and 2 (right), for an initialization with topological obstruction (top) and without it (bottom). The colored curves represent the gradient descent trajectories from initialization $\theta_k(0)$ up to $t_* = 500$ optimization steps. **c.** The loss curves for the bad (obstructed) and good initializations.

Results. We initialize different models and collect all states and losses. First, we initialize the model with an "unlucky" configuration with $c = (-0.1, -0.1)$, i.e. characterized by the space having $2^{l_-} = 4$ connected components (Corollary 3.3). The initialization we chose specifically has $s(\theta) = (+1, +1)$, meaning that the theory would predict the network to be incapable of getting negative output weights. Indeed, we find that the trajectories are confined to the positive region of their invariant hyperquadric, resulting in a poor approximation of f^*, as we can see in Figure 3.3b (top) and in the loss of Figure 3.3c.

Instead, with an initial configuration such that $c = (-0.1, +0.1)$ (2 connected components) and $s(\theta) = (+1, +1)$, the model can leverage the connectedness of $\mathcal{Q}(c_2)$ to learn f^* by flipping the sign of the second neuron's output weight (Figure 3.3b bottom right).

A more realistic experiment. We present here a further experiment to show how the topological obstruction can be a hindrance in a more realistic setting. We consider a simple binary classification task on the well-known *breast cancer* dataset [283], which we try to solve by fitting a one-layer ReLU neural network trained to minimize the BCE loss. We vary the number of hidden neurons l and, for each l, we change the number of non-pathological neurons l_+ (neurons with $c_k > 0$) from 0 to l. We repeat the experiment with 100 different random

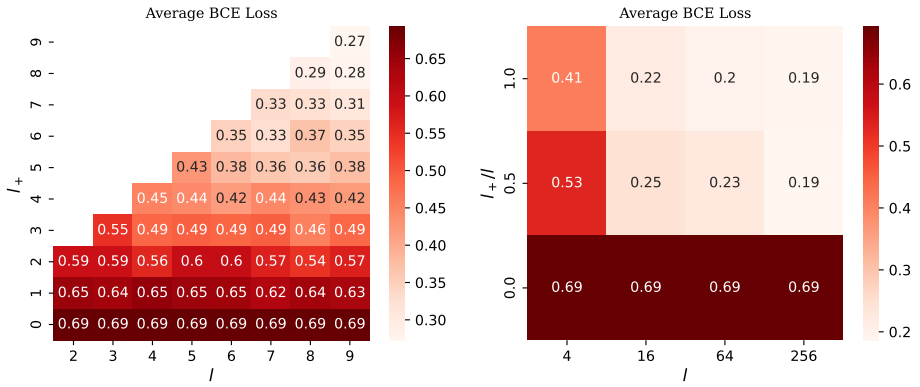

Fig. 3.4 **Left.** Average test BCE loss of a two-layer ReLU neural network trained on the *breast cancer* dataset over 100 different initializations for each pair (l,l_+), $l = 2,\dots,9$ and $l_+ \leq l$, of numbers of hidden neurons and non-pathological neurons. **Right.** the y-axis displays the percentage of non-pathological neurons.

initializations and show how the model's average performance changes when the degree of disconnectedness of its invariant set is varied. The result, on the left panel of Figure 3.4, clearly shows the presence of a "gradient" in performance, where increasing the number of non-pathological neurons decreases the average value of the test loss after training. The right panel of Figure 3.4, moreover, shows how the impact of the obstruction depends on the number of non-pathological neurons and not on their fraction over the total number of hidden neurons.

3.6 Probability of obstruction

Let us consider the following question: *what is the probability of having a disconnected invariant set given a realistic initialization?*

Consider a one-layer ReLU neural network with $e = 1$ and assume that the weights are sampled independently of one another from a normal distribution $W^{(1)}_{ki} \sim \mathcal{N}(0,\sigma_1^2)\ \forall k,i$, $W^{(2)}_{k} \sim \mathcal{N}(0,\sigma_2^2)\ \forall k,j$. From Corollary 3.3, we know that the invariant set $\mathcal{H}(c)$ will be disconnected if and only if there exists a hidden neuron satisfying $\sum_{i=1}^{d}(W^{(1)}_{ki})^2 < (W^{(2)}_{k})^2$. Given independence of the initial weight sampling, this probability can be computed as

$$\mathbb{P}[\text{obstruction}] = 1 - \mathbb{P}\left[\sum_{i=1}^{d}(W^{(1)}_{ki})^2 > (W^{(2)}_{k})^2\right]^l = 1 - (F_{1,d}(d\sigma_1^2/\sigma_2^2))^l,$$

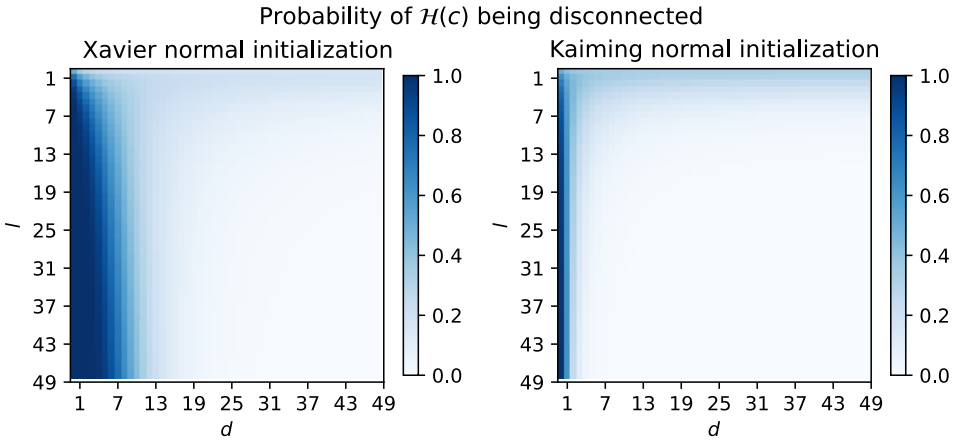

Fig. 3.5 Probability of the topological obstruction as a function of the number of input d and hidden l neurons, when the initial weights are sampled with Xavier normal (left) and Kaiming normal (right) initialization schemes.

where F is the cumulative distribution function of a Fisher-Snedecor-distributed random variable, i.e. given by the ration of two chi-square-distributed random variables.

Having obtained this general expression, we can specify it to two common initialization schemes.

- We obtain *Kaiming initialization* [284] with $\sigma_1^2 = 2/d, \sigma_2^2 = 2/l$ resulting in $\mathbb{P}[\text{obstruction}] = 1 - F_{1,d}(l)^l$.
- We obtain *Xavier normal initialization* [285] with $\sigma_1^2 = 2/(d+l), \sigma_2^2 = 2/(1+l)$ resulting in $\mathbb{P}[\text{obstruction}] = 1 - F_{1,d}(\frac{d+ld}{d+l})^l$.

We plot these two expressions in Figure 3.5. We can see how, for a large number of input neurons d, the probability of obstruction quickly falls to 0 for any number of hidden neurons. Instead, we see an opposite trend for small values of d: the probability of disconnectedness grows with l. Moreover, it is interesting to notice that the region of high obstruction probability is larger for Xavier initialization than for Kaiming initialization, further showing why the latter is preferred when working with ReLU networks.

3.7 Conclusions

In this chapter, we have given analytical results that clarify the nature of the constraints imposed by gradient flow on the parameter space of a two-layer neural network with homogeneous

activations. In the case of a single scalar output, which appears in tasks such as binary classification and scalar regression, we identified initial conditions that lead to a topological obstruction in the form of the parameter space's fragmentation into multiple connected components. This is caused by pathological neurons whose output weights cannot change their sign during training. Moreover, if one also considers the network's symmetries under permutations of the hidden neurons, we find that most of the connected components are equivalent. The number of effective components of the resulting space scales linearly with the number of pathological neurons, contrasting with the exponential growth of the number of connected components obtained without considering the permutation symmetries.

As shown in the last numerical experiment, the lack of non-pathological neurons hinders learning, even when the network's width is scaled. Our probabilistic analysis outlined in Section 3.6, however, shows that with common initialization schemes, the probability of creating a pathological neuron decreases rapidly with increased inner layer width. Therefore, the combination of specific initialization schemes and a large number of hidden neurons (beyond the minimum required to solve a task) appears to make this obstruction unlikely in practice. This chapter describes a simple safeguard to avoid obstructions, which can, for instance, discourage the usage of initialization schemes that result in the proliferation of pathological neurons.

3.8 Limitations

The main limitation of this set of results is the network's architecture, which is limited to only one hidden layer. Even when considering multi-layer neural networks we can define observationa equivalence symmetries in the form of rescalings and permutations, and, as a consequence, find invariant hyperquadrics for each hidden neuron. The issue in extending the results emerges in the fact that these hyperquadrics are *not independent* anymore, and the invariant set cannot be factored into the product of the $Q(c_k)$. This intuitively results from the fact that in the multi-layer case, each weight in the hidden layers is shared by two neurons.

We will see in Chapter 4 how this problem can still be approached by leveraging some recent results about the topology of intersections of hyperquadrics [286]. The new results we obtain will give us an interesting general perspective on the geometrical meaning of gradient flow optimization in a wide class of networks with homogeneous activations.

3.9 Extra proofs and lemmas

Lemma 3.2. *The invariant set $\mathcal{H}(c)$ is connected if and only if it is path connected.*

Proof. Lemma 3.1 tells us that $\mathcal{H}(c) \cong \mathcal{Q}(c_1) \times \cdots \times \mathcal{Q}(c_l)$.

Let us focus on a particular $\mathcal{Q}(c_k)$.

When $c_k \neq 0$, Proposition 3.6 tells us that $\mathcal{Q}(c_k)$ is a topological manifold, and thus, it is connected if and only if it is path connected.

When $c_k = 0$, $\mathcal{Q}(0)$ is not a topological manifold but contractible, implying that it is (path) connected. Let us prove that it is also path-connected.

Let $\theta_k, \theta'_k \in \mathcal{Q}(0)$ and define the curve $\gamma : [0,1] \to \mathcal{Q}(0)$

$$\gamma_{k;\theta}(t) = t \cdot \theta_k$$

such that $\gamma_{k;\theta}(0) = \theta,\ \gamma_{k;\theta}(1) = 0$. $\gamma_{k;\theta}(t) \in \mathcal{Q}(0)$ for every $t \in [0,1]$ because

$$\langle\!\langle \gamma_{k;\theta}(t), \gamma_{k;\theta}(t) \rangle\!\rangle_k = t \langle\!\langle \theta, \theta \rangle\!\rangle_k = 0.$$

Therefore, the segment from θ_k to 0 belongs to $\mathcal{Q}(0)$.

A continuous curve from θ_k to θ'_k can be thus obtained by

$$\gamma_{k;\theta'}\gamma_{k;\theta}(t) := \begin{cases} \gamma_{k;\theta}(2t) & \text{if } t \in [0, \frac{1}{2}] \\ \gamma_{k;\theta'}(2-2t) & \text{if } t \in [\frac{1}{2}, 1] \end{cases}$$

which is continuous because $\gamma_{k;\theta}(1) = \gamma_{k;\theta'}(1) = 0$. Therefore $\mathcal{Q}(0)$ is path connected.

Finally, if $\mathcal{H}(c)$ is connected, then all of its factors $\mathcal{Q}(c_k)$ are connected, which, in turn, is true if and only if they are path-connected. A product of path-connected space is again path-connected, and therefore $\mathcal{H}(c)$ is path-connected. The other implication is true because path-connectedness implies connectedness, thus concluding the proof. □

Lemma 3.3 (Interchange of rescalings and permutations)**.** *Let $\theta \in \Theta$, $\alpha \in \mathbb{R}^l_+$ and $\pi \in \mathcal{S}_l$, then, if $\tilde{\alpha} = R_{\pi^{-1}}(\alpha)$*

$$T_\alpha P_\pi(\theta) = P_\pi T_{\tilde{\alpha}}(\theta). \tag{3.24}$$

Proof. Given that

$$T_\alpha P_\pi(\theta) = (\mathrm{diag}(\alpha) R_\pi W^{(1)}, W^{(2)} R_\pi^\top \mathrm{diag}(\alpha)^{-1})$$

we need to prove that $\mathrm{diag}(\alpha) R_\pi = R_\pi \mathrm{diag}(\tilde{\alpha})$.

$$(\mathrm{diag}(\alpha) R_\pi)_{ij} = \sum_{k=1}^{l} \mathrm{diag}(\alpha)_{ik} (R_\pi)_{kj} = \alpha_i (R_\pi)_{ij} = \begin{cases} \alpha_i & \text{if } j = \pi(i) \\ 0 & \text{otherwise} \end{cases}.$$

Let us pick a generic $\tilde{\alpha} \in \mathbb{R}_+^l$.

$$(R_\pi \mathrm{diag}(\tilde{\alpha}))_{ij} = \sum_{k=1}^{l} (R_\pi)_{ik} \mathrm{diag}(\tilde{\alpha})_{kj} = (R_\pi)_{ij} \tilde{\alpha}_j = \begin{cases} \tilde{\alpha}_j & \text{if } j = \pi(i) \\ 0 & \text{otherwise} \end{cases}.$$

Let us consider the inverse permutation π^{-1} so that $\pi^{-1}(j) = i$ if $\pi(i) = j$. Then, if $\tilde{\alpha} = R_{\pi^{-1}} \alpha$,

$$\tilde{\alpha}_j = \alpha_{\pi^{-1}(j)} = \alpha_i$$

and thus we get that $\mathrm{diag}(\alpha) R_\pi = R_\pi \mathrm{diag}(\tilde{\alpha})$. □

3.9.1 Proof of Proposition 3.8

Proof. We have by Equation (3.11) and Equation (3.2):

$$\langle\!\langle T_\alpha(\theta), T_\alpha(\theta) \rangle\!\rangle_k - c_k = 0 \iff \alpha_k^2 \sum_{i=1}^{d} (W_{ki}^{(1)})^2 - \frac{1}{\alpha_k^2} \sum_{j=1}^{e} (W_{jk}^{(2)})^2 - c_k = 0$$

By renaming $A = \sum_{i=1}^{d} (W_{ki}^{(1)})^2 = \left\| W_k^{(1)} \right\|_F^2$, $C = \sum_{j=1}^{e} (W_{jk}^{(2)})^2 = \left\| W_k^{(2)} \right\|_F^2$ and multiplying by $\alpha_k^2 > 0$ we have:

$$A\alpha_k^4 - c_k \alpha_k^2 - C = 0 \tag{3.25}$$

Solving for α_k^2 gives us:

$$\Delta = c_k^2 + 4AC \geq 4AC > 0$$

$$\alpha_k^2 = \frac{c_k \pm \sqrt{\Delta}}{2A}$$

Given that we want $\alpha_k > 0$, we discard the negative solution. The other is positive because $\Delta > c_k^2$ and thus $\sqrt{\Delta} > |c_k|$.

$$\alpha_k = \pm\sqrt{\frac{c_k + \sqrt{\Delta}}{2A}}$$

Of which we keep the positive solution only, with its full expression being:

$$\alpha_k = \sqrt{\frac{c_k + \sqrt{c_k^2 + 4\left\|W_k^{(1)}\right\|_F^2 \left\|W_k^{(2)}\right\|_F^2}}{2\left\|W_k^{(1)}\right\|_F^2}}. \tag{3.26}$$

Hence, if $\alpha = (\alpha_1, \ldots, \alpha_l)$ with α_k given by Equation (3.26), we get that $\langle\langle T_\alpha(\theta), T_\alpha(\theta)\rangle\rangle_k = c_k \ \forall k = 1, \ldots, l$.

Let us consider now the pathological cases $W_k^{(1)} = 0$ or $W_k^{(2)} = 0$.

If $W_k^{(1)} = 0, W_k^{(2)} \neq 0$ then $A = 0, C \neq 0$. Therefore, we have that Equation (3.25) becomes

$$-c_k\alpha_k^2 - C = 0$$

which has solutions if and only if $c_k < 0$. In that case $\alpha_k = \frac{\left\|W_k^{(2)}\right\|_F}{\sqrt{-c_k}}$. This means that a hidden neuron with zero input weights and nonzero output weights can be rescaled only to the invariant hyperquadrics with $c_k < 0$.

If $W_k^{(2)} = 0, W_k^{(1)} \neq 0$ then $C = 0, A \neq 0$. Therefore, we have that Equation (3.25) becomes

$$\alpha_k^2 A - c_k = 0$$

which has solutions if and only if $c_k > 0$. In that case $\alpha_k = \frac{\sqrt{c_k}}{\left\|W_k^{(1)}\right\|_F}$. This means that a hidden neuron with zero output weights and nonzero input weights can be rescaled only to the invariant hyperquadrics with $c_k > 0$.

If $W_k^{(1)} = 0$ and $W_k^{(2)} = 0$, then $\theta_k = 0 \in \bar{Q}(0)$ and it cannot be rescaled to any other invariant hyperquadric.

□

3.10 Including biases

Let us consider the case where we include biases. The resulting two-layer neural network can be written as

$$f(x;\theta) = W^{(2)}\sigma(W^{(1)}x + b^{(1)}) + b^{(2)}, \tag{3.27}$$

where $b^{(1)} \in \mathbb{R}^l$ and $b^{(2)} \in \mathbb{R}^e$.

To work with this extended set of parameters, we re-define the space

$$\Theta = \left\{\theta = (W^{(1)}, b^{(1)}, W^{(2)}, b^{(2)})\right\}$$

and the single hidden neuron spaces

$$\Theta_k = \left\{\theta_k = (e_{kk}W_k^{(1)}, e_{kk}b_k^{(1)}, W_k^{(2)}e_{kk})\right\}$$

where the second bias term $b^{(2)}$ does not appear because it is not directly involved with the computations of the hidden neurons. This means that we can write

$$\Theta \cong \Theta_1 \times \cdots \times \Theta_l \times \mathbb{R}^e$$

where $\mathbb{R}^e$ is included to describe the parameters in $b^{(2)}$.

The neuron rescaling action now acts on the biases $b^{(1)}$ as well as the weights:

$$\begin{aligned} T\colon \mathbb{R}_+ \times \Theta_k &\to \Theta_k \\ (\alpha, \theta_k) &\mapsto T_\alpha(\theta_k) = (\alpha W_k^{(1)}, \alpha b_k^{(1)}, \frac{1}{\alpha} W_k^{(2)}) \end{aligned} \tag{3.28}$$

and can be extended to the whole space of parameters

$$\begin{array}{rccl} T\colon & \mathbb{R}_+^l \times \Theta & \to & \Theta \\ & (\alpha, \theta) & \mapsto & T_\alpha(\theta) = (\mathrm{diag}(\alpha)W^{(1)}, \mathrm{diag}(\alpha)b^{(1)}, W^{(2)}\,\mathrm{diag}(\alpha)^{-1}, b^{(2)}). \end{array} \tag{3.29}$$

Once again, we find that $T_\alpha\theta \sim \theta$.

In this more general case, we can rewrite the bilinear form to include the biases. If $\theta = (W^{(1)}, b^{(1)}, W^{(2)}, b^{(2)})$ and $\eta = (V^{(1)}, p^{(1)}, V^{(2)}, p^{(2)})$, we define

$$\langle\langle \theta, \eta \rangle\rangle_k = \sum_{i=1}^{d} W_{ki}^{(1)} V_{ki}^{(1)} + b_k^{(1)} p_k^{(1)} - \sum_{j=1}^{e} W_{jk}^{(2)} V_{jk}^{(2)} \tag{3.30}$$

and see that, once gradient flow optimization, we have a conservation condition like the one of Equation (3.8)

$$\langle\langle \theta(t), \theta(t) \rangle\rangle_k = c_k \forall t > 0 \; \forall k = 1, \dots, l.$$

Once again, we call $\mathcal{Q}(c_k)$ the hypersurface of Θ_k which satisfies the equation $\langle\langle \theta, \theta \rangle\rangle_k = c_k$ and $\mathcal{H}(c_k)$ the set in Θ defined by $\langle\langle \theta, \theta \rangle\rangle_k = c_k \; \forall k = 1, \dots, l$.

With this in mind, it is not hard to extend the results of Proposition 3.6 and Corollary 3.1 which turn out to be slightly modified.

Proposition 3.9. *If $c_k > 0$, $\mathcal{Q}(c_k)$ is a topological manifold homeomorphic to $\mathbb{R}^e \times S^d$. If $c_k < 0$, $\mathcal{Q}(c_k)$ is a topological manifold homeomorphic to $\mathbb{R}^d \times S^{e-1}$. If $c_k = 0$, $\mathcal{Q}(0)$ is contractible.*

In this case, we can factor the space of parameters $\mathcal{H}(c)$ as

$$\mathcal{H}(c) \cong \mathcal{Q}(c_1) \times \cdots \times \mathcal{Q}(c_l) \times \mathbb{R}^e,$$

where the last factor is due to the freedom in choosing the values of $b^{(2)}$.

Proposition 3.10. *Let $c_k \neq 0 \; \forall k = 1, \dots, l$. Let l_+, l_-, l_0 be the number of positive, negative and zero elements of c, respectively. The Poincaré polynomial of $\mathcal{H}(c)$ is given by*

$$p_{\mathcal{H}(c)}(x) = (1 + x^d)^{l_+} (1 + x^{e-1})^{l_-} \tag{3.31}$$

Corollary 3.5. *The 0-th Betti number $\beta_0(c)$ of $\mathcal{H}(c)$, corresponding to the number of its connected components, is given by*

$$\beta_0(c) = \begin{cases} 1 & \text{if } e > 1 \\ 2^{l_-} & \text{if } e = 1 \end{cases} \tag{3.32}$$

The result we obtain is similar to Corollary 3.2 although slightly modified by the fact that having a single input neuron does not cause $\mathcal{H}(c)$ to become disconnected anymore. In the case of $e = 1$, therefore, the picture presented in the main text is left unchanged.

Back to Table of Contents

Chapter 4

Topology and Geometry of the learning space of ReLU networks

Contents

4.1 Introduction

Building upon the topological foundations established in Chapter 3, we now extend our framework to a more general class of neural networks. While our earlier analysis was limited to shallow architectures, this chapter generalizes these concepts to feed-forward networks defined by arbitrary directed acyclic graphs (DAGs). This shift allows us to move beyond the simple

two-layer structure and investigate how multilayered and more complex connectivity patterns can influence the geometry of the optimization space. We demonstrate that the *invariant set* remains a central object of study, but its algebraic structure becomes significantly more intricate as the network depth and complexity increase.

A major contribution of this chapter is the dual focus on both the global connectivity of the parameter space and its local geometric pathologies, specifically *singularities*. By leveraging tools from algebraic topology and network flow theory, we characterize the singular points of the invariant set—configurations where the local dimensionality of the space collapses. We show that these singularities are not merely mathematical peculiarities but are deeply linked to the architectural properties of the DAG, specifically to configurations where parts of the network become functionally disconnected.

Finally, we bridge the gap between this geometric theory and practical model compression. We introduce a principled approach to sparsity by showing how specific regularizers, such as the nuclear norm of the invariant set's Jacobian, can be used to drive the optimization trajectory toward these singular configurations. This perspective allows us to frame pruning not as a heuristic process of removing small weights, but as a geometrically grounded transition toward singular subnetworks.

4.2 Setup

We start by extending the structure of two-layer neural networks outlined in Section 3.2.1, by defining more general neural networks whose architecture is given by a directed acyclic graph.

DAG neural networks. A general **DAG neural network** is a directed, acyclic graph G (i.e. a DAG) on a set of nodes V, called neurons, with edges E.

We identify a subset of neurons $V_I \subseteq V$ containing the **input neurons**, such that no edges are entering the elements of V_I, and a subset $V_O \subseteq V$ of **output neurons** such that no edges are going out of the elements in V_O. We assume that $V_I \cap V_O = \emptyset$, i.e., no neurons have empty neighbors. We write ∂V to denote the set of input and output neurons $V_I \sqcup V_O$. All the other nodes $\tilde{V} \subseteq V$ are the **hidden neurons** which are the fundamental computational units of the neural network $V = \tilde{V} \sqcup \partial V$ (Figure 4.1a).

For any node $v \in V$, we call $\mathrm{Anc}(v)$ the set of its **ancestors**, i.e. nodes $w \in V$ such that there exists a directed path in G from w to v, and $\mathrm{Desc}(v)$ its **descendants**, i.e. the nodes $w \in V$ such that there exists a directed path from v to w in G.

Each edge $(i, j) \in E$ has a parameter $\theta_{(i,j)} \in \mathbb{R}$ associated with it and, just like in the two-layer case, the computation of the network proceeds as follows:

1. data is fed into the network through the input neurons, which perform no computation;
2. each input is transmitted through the edges coming out of the input neurons, where it is multiplied by each edge's parameter $\theta_{(i,j)}$;
3. every hidden neuron $v \in \tilde{V}$ sums the values of their incoming edges, applies the activation function σ and transmits to each outgoing edge the resulting value multiplied by the edge parameter;
4. the computation stops at the output neurons, where the neural network's output is located.

The fact that G admits no directed cycles ensures that G admits a *topological ordering*, i.e. an ordering of the hidden neurons $\tilde{V}$ such that if $(u, v) \in E$ is a directed edge, then v will come after u in the ordering. This means that the computation of the network can be neatly organized into a finite sequence of steps, without infinite feedback loops. We write $f_G(\cdot, \theta)$ to indicate the input-output function encoded by G with parameters θ and resulting from the computation outlined above.

We call **parameter space** Θ the set of all parameters and interpret it as the vector space of real functions over the edges $\theta : E \to \mathbb{R}$, $\Theta \cong \mathbb{R}^{|E|}$. For this reason, we can see the parameter space as the order 1 real cochain space of G, $\Theta = C^1(G; \mathbb{R})$ (Definition 0.14), if we think of G as an oriented simplicial complex with edge orientations given by their direction.

Remark 4.1 (Two kinds of topology)**.** *The perspective of the parameter space as the chain space will prove to be useful to derive some of the results of this chapter. In particular, it will allow us to talk about how the two "kinds" of topology implicit in this setup are related to one another: the* discrete topology of the computational graph *G, seen as a oriented simplicial complex studied with the tools of simplicial (relative) homology, and the* continuous topology of the invariant set *that constrains the learning trajectories, studied with singular homology.*

This is already implicit in the results of Chapter 3 that, although limited to the simple case of two-layer architectures, tells us how the topology of the invariant set depends on architectural parameters *of the graph, i.e. the number of input, output and hidden neurons.*

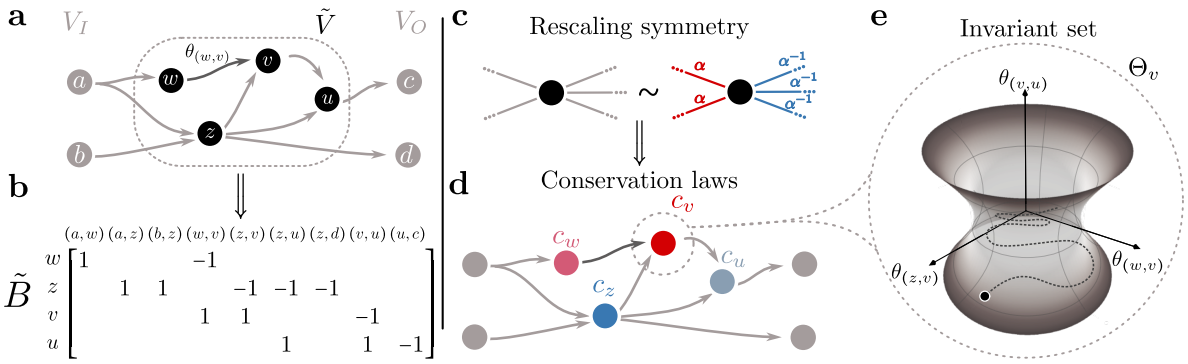

Fig. 4.1 **a.** Example of a feed-forward DAG architecture G. **b.** The incidence matrix $\tilde{B}$ of G with rows associated to input and output neurons removed. **c.** Visualization of the rescaling symmetry of ReLU neurons. **d.** The initialization determines the balance value $c_v = \langle \theta, \theta \rangle_v$ of every hidden neuron, which characterizes the shape of the invariant set (**e**).

Throughout the chapter, it will be convenient to formulate the results by describing the graph connectivity structure with the node-edge incidence matrix B of G, introduced in Definition 0.5. We recall that its elements are defined as follows: $B_{v,(u,w)} = 1$ if $v = w$, $B_{v,(u,w)} = -1$ if $v = u$ and 0 otherwise. See, for example, the DAG in Figure 4.1a and its associated incidence matrix (with rows associated to nodes in ∂V removed) in Figure 4.1b.

Symmetries of DAG ReLU networks. Just as we did in Chapter 3, we study the properties of neural network where σ is a *homogeneous* activation function, like the ReLU $\sigma(x) = \max(0, x)$.

We recall from Chapter 3, that the rescaling action invariance we studied for two-layer ReLU networks (Equation (3.2)) does not depend on the two-layer structure of the network but is a "local" property that holds independently at every hidden neuron. It follows that it holds in the same way in general DAG architectures (Figure 4.1c). We consider the action of the group $\mathbb{R}_+$ of positive real numbers on each local parameter space of hidden neuron $v \in \tilde{V}$, $\Theta_v := \{\theta_{(u,w)} \mid (u, w) \in E \text{ and } u = v \text{ or } w = v\}$ by means of

$$T^v_\alpha(\theta) = T^v_\alpha((\theta_{(u,v)})_u, (\theta_{(v,w)})_w) = ((\alpha\theta_{(u,v)})_u, (\alpha^{-1}\theta_{(v,w)})_w).$$

The invariance of the neural network function to the rescaling is then written in the following way.

Proposition 4.1 (Local rescaling invariance)**.** *Let $G = (V, E)$ be a DAG neural network. We have that θ and $T^v_\alpha(\theta)$ are observationally equivalent parameters*

$$f_G(\cdot; \theta) = f_G(\cdot; T^v_\alpha(\theta)),$$

for every $\theta \in \Theta$, $v \in \tilde{V}$ *and* $\alpha \in \mathbb{R}_+$.

Proof. Consider the contribution of v to the neural network processing when its parameters are rescaled through T^v_α. If $\{x_w | (w, v) \in E\}$ is the set of activation coming into v through its incoming connections, any neurons z that receive an input from v will receive the value

$$x_z = \theta_{(v,z)} \sigma \left(\sum_{w:(w,v)\in E} \theta_{(w,v)} x_w \right),$$

which is invariant when $\theta_{(v,z)} \mapsto \frac{1}{\alpha}\theta_{(v,z)}$ and $\theta_{(w,v)} \mapsto \alpha\theta_{(w,v)}$. □

Local conservation laws under gradient flow. We pick a dataset $D = \left\{(x_i, y_i) \in \mathbb{R}^d \times \mathbb{R}^e\right\}_{i=1}^N$ and, just like in Equation (3.5), a loss function $L : \Theta \to \mathbb{R}$ that depends on the parameters only through the output of the neural network[1], that is

$$L(\theta) = \frac{1}{N} \sum_{i=1}^{N} \ell(f_G(x_i; \theta), y_i) \tag{4.1}$$

where $\ell : \mathbb{R}^{|V_O|} \times \mathbb{R}^{|V_O|} \to \mathbb{R}$ is differentiable.

Once again, we train the network using gradient flow $\dot{\theta}(t) \in -\nabla_\theta L(\theta(t)) := -g(\theta(t))$. Given that the loss function L depends on the parameters only through f, its value at θ must be constant over the orbit of rescaling. Retracing the proof of Proposition 3.4 using the action T^v_α, leads us to the following generalized balance equation.

$$\langle\!\langle \theta, g(\theta) \rangle\!\rangle_v := \sum_{u:(u,v)\in E} \theta_{(u,v)} g_{(u,v)} - \sum_{w:(v,w)\in E} \theta_{(v,w)} g_{(v,w)} = 0, \tag{4.2}$$

for every hidden neuron $v \in \tilde{V}$. Note that in Equation (4.2) we defined the balance form $\langle\!\langle \cdot, \cdot \rangle\!\rangle_v$ associated to a hidden neuron v in a DAG neural network, which generalizes the two-layer balance form $\langle\!\langle \cdot, \cdot \rangle\!\rangle_k$ of Definition 3.2.

Remark 4.2. *Equation (4.2) tells us that the gradient modulated by the parameter values is a* ***network flow****, that is the quantity* $g(\theta) \odot \theta$ *is conserved when passing through each hidden neuron, with* $\odot$ *denoting the element-wise Hadamard product between vectors.*

[1] This means that we do not include regularization terms which depend explicitly on the parameters.

It turns out that we can conveniently re-write the balance conditions of Equation (4.2) at all hidden nodes using the incidence matrix of G.

Proposition 4.2 (Matrix rewriting of the balance equations)**.** *Let $\tilde{B} \in \mathbb{R}^{|\tilde{V}|\times|E|}$ be the incidence matrix of G with the rows associated with input and output nodes removed; then $\langle\langle \theta, g(\theta)\rangle\rangle_v = 0\ \forall v \in \tilde{V}$ is equivalent to*

$$\tilde{B}(\theta \odot g(\theta)) = 0. \tag{4.3}$$

Proof. The proof follows directly from the definition of the incidence matrix. At any hidden node $v \in \tilde{V}$, if we denote by θ_e the weight of any $e \in E$, and by g_e the e-th component of $g(\theta)$, then it holds that

$$\tilde{B}(\theta \odot g(\theta))_v = \sum_{e\in E} B_{v,e}\theta_e g_e = \sum_{u:(u,v)\in E} \theta_{(u,v)} g_{(u,v)} - \sum_{w:(v,w)} \theta_{(v,w)} g_{(v,w)} = \langle\langle \theta, g(\theta)\rangle\rangle_v = 0.$$

□

Remark 4.3 (Balance equation and relative homology)**.** *Elaborating on the discussion we started in Remark 4.1, note that, in the generalized gradient balance condition of Proposition 4.2, the matrix $\tilde{B}$ appears. As we write in the proposition's text, this is* not *simply the incidence matrix of the underlying computational graph G because the rows corresponding to input and output nodes are removed. The reason for this, on one hand, is simple because input and output neurons do not apply the activation function and thus do not have balance equations associated to them.*

On the other, it turns out that the structure of $\tilde{B}$ can be neatly interpreted under the light of relative homology (Definition 0.13). The reason for this is explained in detail in Example 0.4 in the preliminaries chapter, but can be summarized by saying that $\tilde{B}$ describes node-edge incidence relations of the graph G modulo *nodes that belong to the input or output set $\partial V = V_I \sqcup V_O$. Therefore, Proposition 4.2 tells us that we can interpret $\theta \odot g(\theta)$ as a representative of the* 1*-st relative homology class of the pair $(G, \partial V)$. If we put this together with the characterization of Proposition 0.2 (as we explain in Example 0.4), this means that $\theta \odot g(\theta)$ is a cycle on the graph G with all the nodes in ∂V collapsed to a single node.*

Invariant sets. Just like Proposition 3.4 in the two-layer case, Proposition 4.2 implies that some quantities are conserved under gradient flow optimization or, equivalently, that the learning trajectories are constrained to a lower-dimensional subset of the parameter space.

Proposition 4.3 (DAG conservation laws)**.** *Let the network be initialized with parameters* $\theta(0)$ *such that* $\tilde{B}\theta(0)^2 = c \in \mathbb{R}^{|\tilde{V}|}$*, then, for every* $t \geq 0$*, it holds that* $\tilde{B}\theta(t)^2 = c$ *for every* $t \geq 0$*, where* $\theta(t)^2$ *is the component-wise square of* $\theta(t)$*.*

Proof.

$$\frac{\mathrm{d}}{\mathrm{d}t}\tilde{B}\theta(t)^2 = \tilde{B}\frac{\mathrm{d}}{\mathrm{d}t}\theta(t)^2 = 2\tilde{B}(\theta(t) \odot \dot{\theta}(t)) = -2\tilde{B}(\theta(t) \odot g(\theta(t)) = 0.$$

□

As we will see, this is not a mere notational feature, as this formulation reveals precious insights into the relation between the training dynamics and the neural network's graph structure. We now define the invariant set given by these new, generalized conservation laws.

Definition 4.1 (Invariant set)**.**

Let $\tilde{B} \in \mathbb{R}^{|\tilde{V}|\times|E|}$ be the incidence matrix of G with the rows associated with input and output nodes removed. Given $c \in \mathbb{R}^{|\tilde{V}|}$, we call *invariant set* the set $\mathcal{H}_G(c) \subseteq \Theta$ of the solutions of the system of polynomial equations $\tilde{B}\theta^2 = c$.

If we look at the single equation associated with the hidden neuron $v \in \tilde{V}$, we see that $\tilde{B}\theta^2 = c$ can be written as

$$\sum_{u:u\to v} \theta^2_{(u,v)} - \sum_{w:w\leftarrow v} \theta^2_{(v,w)} = c_v \tag{4.4}$$

which corresponds to a hyperbolic *quadric hypersurface* in the local parameter space of v, Θ_v (Figure 4.1e). From the graph's point of view, we can interpret Equation (4.4) as stating that the vector of squared parameters θ^2 is akin to a fluid flowing through the edges of G, with input/output nodes acting as unconstrained sources/sinks and hidden nodes supplying or demanding some flow according to the value and sign of c [287].

In Chapter 3, we showed that for shallow networks, the geometrical structure of $\mathcal{H}_G(c)$ is simple, as the total invariant set factorizes into the Cartesian product of the neurons' hyperquadrics. In the general case (MLP or DAG), the situation is much more complex because the equations of $\mathcal{H}_G(c)$ are *coupled*: parameters associated with internal edges appear in multiple equations.

Example 4.1 (Two-layer)**.** *For a two-layer neural network with d input neurons, l hidden neurons and e output neurons, we can write the incidence matrix as*

$$\tilde{B} = \begin{matrix} \\ \textit{Hidden} \end{matrix} \overset{\textit{Input-hidden}}{\left[\begin{array}{ccc|ccc} I_l & \cdots & I_l & -I_l & \cdots & -I_l \end{array}\right]} \overset{\textit{Hidden-output}}{}.$$

We see that in every column, corresponding to one edge, there is a single nonzero element, i.e. the one associated to the hidden neuron it is connected to. This means that every parameter θ_e appears in only one equation of $\tilde{B}\theta^2 = c$.

The invariant set $\mathcal{H}_G(c)$, which is an *algebraic variety*, is an interesting object because it sits in between the redundant but more "concrete" parameter space and the abstract function space (or *neuromanifold* [288, 289]) the model's implemented function lives in. In fact, fixed any c, no two parameters in $\mathcal{H}_G(c)$ are observationally equivalent w.r.t. rescalings, that is $f_G(\cdot;\theta) = f_G(\cdot;T^v_\alpha(\theta))$ for no rescaling, thus making the invariant set a good proxy for the function space $\mathcal{F}_G := \{f_G(\cdot;\theta) \,|\, \theta \in \Theta\}$. Nevertheless, as discussed in Chapter 3, different values of c can correspond to different topologies of $\mathcal{H}_G(c)$, meaning that $\mathcal{F}_G$ does not provide the full picture to understand the learning process.

4.3 Geometry and topology of the invariant set

$\mathcal{H}_G(c)$ is the set of solutions of a system of degree-two polynomial equations, each one corresponding to a quadric. Despite the apparent simplicity, studying general intersections of quadrics is not easy [286] but, in our case, the specific structure of $\mathcal{H}_G(c)$ greatly simplifies the process. In fact, the equations $\tilde{B}\theta^2 = c$ correspond to a system of *coaxial* quadric hypersurfaces, meaning that they contain only squares and no mixed terms of the form $\theta_e\theta_{e'}$. This fact allows us to employ the powerful recent results obtained in Ref. [286] related to the intersection of coaxial quadrics.

4.3.1 Non-emptiness

The first result concerns the non-emptiness of $\mathcal{H}_G(c)$ for a given c or, from the other point of view, the set of possible balance values c that can result from an initialization.

Proposition 4.4 (Feasible balance)**.** *Assume that for every hidden neuron $v \in \tilde{V}$, there exists a directed path in G connecting the input neurons to v and a directed path connecting v to the output neurons. Then, for all $c \in \mathbb{R}^{|\tilde{V}|}$, one has $\mathcal{H}_G(c) \neq \emptyset$.*

This result means that any balance configuration on the hidden neurons c is achievable through a parameter initialization provided that all neurons are connected to input and output, i.e. they take part in the network's computations. From the point of view of network flows, this tells us that it is possible to build a flow θ^2 that satisfies any supply and demand on the hidden nodes.

The proof, we see, builds upon the fact that the input and output nodes are not involved in the balance equation, allowing us to manipulate the edges coming out of them more "freely" to satisfy the constraint.

Proof. To show that, for any $c \in \mathbb{R}^{|\tilde{V}|}$, the system $\tilde{B}\theta^2 = c$ admits a non-negative solution θ^2, we take a constructive approach and build $x = \theta^2$ explicitly.

Let us initialize $x_0 = \mathbb{1}$ by assigning the value 1 to every edge $(x_0)_e = 1 \ \forall e \in E$. Let us define the balance vector at initialization, $c_0 = \tilde{B}\mathbb{1}$ which, in general, will be different from c, $c_0 \neq c$. For each hidden neuron $v \in \tilde{V}$, we make an adjustment to x_0 that corrects the balance value at v.

If $(c_0)_v < c_v$, define $\delta_v = c_v - (c_0)_v$. Let us consider a path p_v going from the input neurons V_I to v (which exists by definition of G) and the vector $\mathbb{1}_{p_v} \in \mathbb{R}^{|E|}$ supported on it, that is such that it assigns value 1 to each edge in p_v and 0 to all other edges. Consider now $x' = x_0 + \delta_v \mathbb{1}_{p_v}$, which is non-negative because $\delta_v \geq 0$. The balance $c'_w := (\tilde{B}x')_w$ will remain unchanged in all nodes w different from v, $c'_w = (c_0)_w$ because, in such nodes, the same quantity is added to one edge coming in and one edge going out. In v, the new balance value will be

$$c'_v = (\tilde{B}x')_v = (\tilde{B}x_0)_v + (\tilde{B}\mathbb{1}_{p_v})_v = (c_0)_v + \delta_v = (c_0)_v + c_v - (c_0)_v = c_v.$$

If instead $(c_0)_v > c_v$, we define $\delta_v = (c_0)_v - c_v$ and add the vector $\mathbb{1}_{p_v}$ supported on a path p_v that connects v to the output neurons V_O. Once again, in all nodes except for v the balance is left unchanged, while in v

$$c'_v = (\tilde{B}x')_v = (\tilde{B}x_0)_v + (\tilde{B}\mathbb{1}_{p_v})_v = (c_0)_v - \delta_v = c'_v - ((c_0)_v - c_v) = c_v.$$

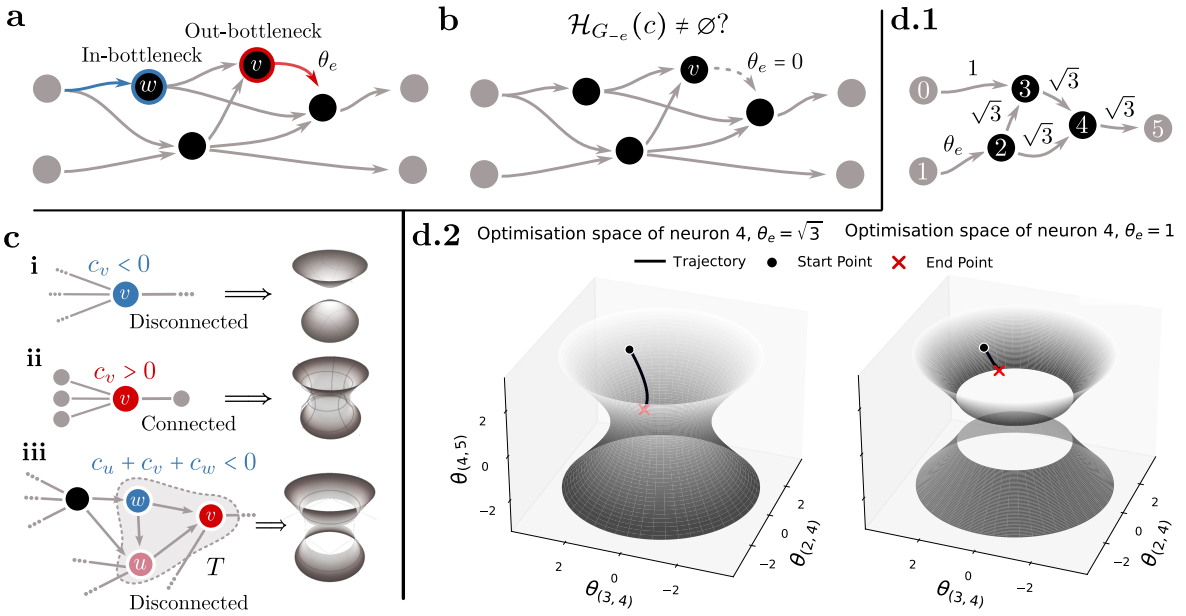

Fig. 4.2 **Overview of connectedness. a.** In- and out-bottlenecks in G. **b.** The non-emptiness of $\mathcal{H}_G(c)$ is guaranteed if every hidden neuron has input and output edges. **c.** Different connectedness conditions and intuitive visualizations of the associated algebraic varieties for an out-bottleneck **d.** Numerical experiment showcasing training dynamics in a connected and disconnected scenario for a DAG network with 3 hidden nodes (**d**.1).

Repeating this process for any node $v \in \tilde{V}$ allows us to correct the balance $c'_v = c_v$ at all nodes, giving us a non-negative solution x' to the system $\tilde{B}x' = c$. □

Example 4.2 (Counterexample to Proposition 4.4)**.** *If the assumption of Proposition 4.4 is not satisfied, it is possible to find balance configurations $c \in \mathbb{R}^{|\tilde{V}|}$ that are not feasible. We take, as a simple example, the case where there is a hidden neuron v with no outgoing edges $\deg^+(v) = 0$. This means that there exists no path from v to V_O, breaking the assumption off the proposition. In this case, any $c \in \mathbb{R}^{|\tilde{V}|}$ such that $c_v < 0$ will be unfeasible, because the balance equation Equation (4.2) at v will be*

$$\sum_{u:u\to v} \theta^2_{(u,v)} = c_v < 0,$$

which admits no real solution θ.

4.3.2 Connectedness

In Chapter 3, we showed that, for some values of c, the invariant set of a shallow ReLU neural network is disconnected. Here, we show that the conditions for connectedness in the general DAG case resemble the ones for the shallow case, with additional pathological cases resulting from particular graph topologies.

To find out whether $\mathcal{H}_G(c)$ is connected or not, we will use the following proposition, taken from Ref. [286].

Proposition 4.5 ([286] Proposition 4.7)**.** *Given a matrix $C = [C_1 \cdots C_n]$ with columns C_i and a vector b, regarding the intersection of coaxial quadrics $X = \{x | Cx^2 = b\}$, the following are equivalent:*

1. *X is connected.*
2. *For every i, $X \cap \{x | x_i = 0\}$ is nonempty.*
3. *b belongs to the convex cone of any $n-1$ of the columns of C, $b \in \mathrm{Co}(C_{-i})$, with $C_{-i} := [C_1 \cdots C_{i-1} C_{i+1} \cdots C_n]$ for all $i = 1, \ldots, n$.*

This proposition tells us that an intersection of quadrics is connected if and only if it is nonempty when sliced with any hyperplane of the form $\{x | x_i = 0\}$. If $C = [C_1 \cdots C_n]$ with columns C_i, then we see that, if $x_i = 0$, $Cx^2 = [C_1 \cdots C_{i-1} C_{i+1} \cdots C_n] x_{-i}^2$. This means that $X \cap \{x | x_i = 0\}$ is equivalent to the intersection of quadrics specified by the matrix C with the i-th column removed and the same b, and thus its nonemptiness can be studied with Proposition 4.4.

In our case $C = \tilde{B}$. Dropping one of its columns is therefore equivalent to *cutting* one of the edges of G. It follows that we can translate Proposition 4.5 to our case in the following way.

Proposition 4.6 (Connectedness as robustness)**.** *$\mathcal{H}_G(c)$ is connected if and only if, for every $e \in E$, $\mathcal{H}_{G_{-e}}(c) \neq \emptyset$, where $G_{-e} = (V, E \setminus \{e\})$.*

In other words, $\mathcal{H}_G(c)$ is connected if it is "robust enough" so that the deletion of single edges doesn't change the possibility of satisfying the supply and demand conditions of c. This observation, together with Proposition 4.4 telling us that supplies and demands are always satisfied when every neuron is connected to V_I and V_O, suggests that disconnection must necessarily come from the presence of neurons with only one input or output connection, so that cutting it *can* make the invariant set empty.

Definition 4.2 (Bottleneck neurons)**.**

A hidden neuron $v \in \tilde{V}$ is an **in-bottleneck** if $\deg^-(v) = 1$ and an **out-bottleneck** if $\deg^+(v) = 1$. We denote with V_B^-, V_B^+ the sets of in and out-bottleneck nodes, respectively. For any out-bottleneck neuron $v \in V_B^+$, we call $\overline{\text{Anc}}(v)$ the set of its *pure ancestors*, i.e. ancestors $w \in \text{Anc}(v)$ such that any path from w to V_O passes through v. Analogously, any in-bottleneck defines a set of *pure descendants* $\overline{\text{Desc}}(v)$ containing neurons $w \in \text{Desc}(v)$ such that any path from V_I to w passes through v.

The removal of the single connection of a bottleneck neuron (Figure 4.2a,b) will disconnect it and the set of its pure ancestors/descendants will be effectively cut out from the network's computation: either because they receive no inputs from V_I or because they produce no output to V_O.

Armed with these tools, it turns out that it is possible to derive a complete characterization of connectedness and disconnectedness.

Theorem 4.1 (Connectedness of the invariant set)**.** *$\mathcal{H}_G(c)$ is connected if and only if*

1. *For all $v \in V_B^+$ and for all $T \subset \overline{\text{Anc}}(v)$ such that T is closed by forward edges $\sum_{u \in T} c_u \geq 0$;*
2. *For all $v \in V_B^-$ and for all $T \subset \overline{\text{Desc}}(v)$ such that T is closed by backward edges $\sum_{u \in T} c_u \leq 0$.*

Before getting into the proof, let us discuss the result and its implications.

Intuitively, disconnectedness is caused by bottleneck neurons such that cutting their single edge makes the balances (supplies/demands) of their pure ancestors/descendants unfeasible. For instance, let us look at the out-bottleneck v in Figure 4.2c:

i) $c_v < 0$ means that v requires more flow coming out than in. This is not feasible because, upon cutting the single output connection, there is no way for flow to come out of v.

ii) $c_v > 0$ means that v requires more flow coming in than out. This is feasible unless

iii) there is set T of neurons closed by forward edges for which $\sum_{k \in T} c_k < 0$, as predicted by Theorem 4.1. Intuitively, this means there is more flow passing through T than can be absorbed before v.

Two immediate corollaries follow, clarifying the problem of connectedness in most practical cases. The first identifies that all networks that are "robust enough" to the deletion of single edges always lead to connected invariant sets.

Corollary 4.1. *If G has no bottleneck neurons, then $\mathcal{H}_G(c)$ is connected.*

Note that bottleneck neurons are rare in MLPs because their existence implies the presence of a layer with only one neuron. As we observed with two-layered architectures, the exceptions are given by neural networks built to solve binary classification or scalar regression problems, where there is a single output neuron, and thus the neurons in the last hidden layer are all out-bottlenecks. In that case, we adapt Theorem 4.1, finding that the result of Corollary 3.3 nicely carries over to the multi-layer case.

Corollary 4.2 (The MLP case)**.** *If G is a fully-connected MLP such that all hidden layers contain more than one neuron, then $\mathcal{H}_G(c)$ is connected if and only if $c_v \geq 0 \; \forall v \in V_B^+$ and $c_v \leq 0 \; \forall v \in V_B^-$.*

Proof. The result follows from Theorem 4.1 by noticing that in a fully connected architecture, the only pure ancestor/pure descendant of a node is the node itself. □

A practical implication of this section is that the expressivity of ReLU networks can be reduced at initialization to the extent that they lose their universal approximation capability. That is, some functions become immediately unreachable, regardless of the chosen loss function or dataset.

Numerical validation. We illustrate Theorem 4.1 on the toy DAG neural network shown in Figure 4.2d.1, implemented using dedicated software [290] and trained with discrete gradient descent. Neuron 4 will be the out-bottleneck of interest. All hidden neurons (in black) have ReLU activation functions. With the initialization given by the values in the figure, we have $c = (c_2, c_3, c_4) = (\theta_e^2 - 6, 1, 3)$. There are 3 forward stable sets of nodes, the largest being $T = \{2, 3, 4\}$ with $\sum_{k\in T} c_k = \sum_k c_k = \theta_e^2 - 4$. Therefore, if at initialization $\theta_e(0) < \sqrt{2}$, then $\sum_{k\in T} c_k < 0$ and the optimization space will disconnect at neuron 4. Concretely, it means that the balance condition at neuron 4 will forbid sign switches of $\theta_{(4,5)}$.

We try to make the model learn the function $f : (x_1, x_2) \to -(x_1 + x_2)$ for positive inputs. The only way to output negative values is if $\theta_{(4,5)} < 0$, and so the optimum will not be reachable for $\theta_e(0) = 1 < \sqrt{2}$ as shown on the right plot of Figure 4.2d.2, while the optimum is reachable (and reached) for $\theta_e(0) = \sqrt{3} > \sqrt{2}$, as depicted on the left plot.

Proof of Theorem 4.1. The proof proceeds in three steps. First, we associate to each bottleneck node a *network flow feasibility problem*. Second, we show that connectedness of $\mathcal{H}_G(c)$ is equivalent to feasibility of all these flow problems. Finally, we characterize feasibility via cut inequalities.

Step 1: Flow problems associated to bottlenecks. Let $v \in V_B^+$ be an out-bottleneck. We construct a directed graph $G_v = (N_v, A_v)$ as follows.

The node set is

$$N_v := \overline{\mathrm{Anc}}(v) \sqcup \{s, t\},$$

where $\overline{\mathrm{Anc}}(v)$ denotes the set of pure ancestors of v and s, t are auxiliary source and sink nodes.

The arc set A_v consists of:

- *Internal arcs* (u, w) for every edge $(u, w) \in E$ with $u, w \in \overline{\mathrm{Anc}}(v)$.
- *Incoming arcs* (s, w) for every edge $(u, w) \in E$ with $u \notin \overline{\mathrm{Anc}}(v)$ and $w \in \overline{\mathrm{Anc}}(v)$.
- *Source arcs* (s, u) for each $u \in \overline{\mathrm{Anc}}(v)$ with $c_u < 0$.
- *Sink arcs* (u, t) for each $u \in \overline{\mathrm{Anc}}(v)$ with $c_u > 0$.
- A *circulation arc* (t, s).

Note that there are no outgoing arcs from pure ancestors to non pure ancestors as $\overline{\mathrm{Anc}}(v)$ is closed by taking forward edges. In fact, if $(u, w) \in E$ with $u \in \overline{\mathrm{Anc}}(v)$, $w \notin \overline{\mathrm{Anc}}(v)$, then, by definition, there would be a path from w to the outputs V_O not passing through v. Adding the edge (u, w) to that path would give us a path from u to V_O that doesn't pass from v, meaning that $v \notin \overline{\mathrm{Anc}}(v)$.

Each arc e carries a flow $f_e \in \mathbb{R}$ subject to the following constraint:

$$\begin{cases} 0 \leq f_e < \infty & e \text{ internal, incoming, or } (t, s), \\ f_{(s,u)} = -c_u & c_u < 0, \\ f_{(u,t)} = c_u & c_u > 0. \end{cases} \tag{4.5}$$

The network (N_v, A_v) is said to be *feasible* if there exists a real function on the edges, called *flow* and denoted $f : A_v \to \mathbb{R}$ such that

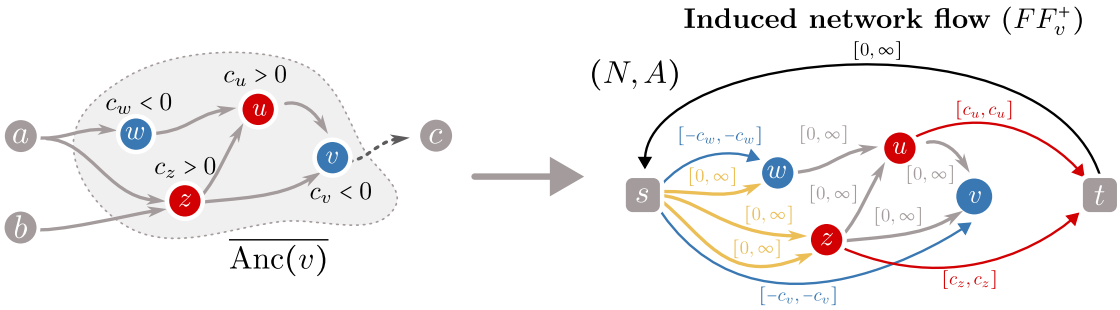

Fig. 4.3 Visualization of the induced network flow problem (FF_v^+) at a bottleneck node $v \in V_B^+$. In the right panel, we depict the internal arcs in gray, the incoming arcs in orange, the source arcs in blue, the sink arcs in red and the circulation arc in black.

1. at each node the flow is conserved $\sum_{u:(u,w)\in A_v} f_{(u,w)} - \sum_{z:(w,z)\in A} f_{(w,z)} = 0 \; \forall w \in N$;
2. the constraints are respected on each edge $l_e \leq f_e \leq m_e \; \forall e \in A_v$.

We denote this *flow feasibility problem* on neuron v by (FF_v^+), which is said to be in *circulation form* due to the existence of the circulation arc (t, s) [291]. The analogous problem (FF_v^-) for an in-bottleneck is defined by reversing all arrows and replacing c with $-c$.

Step 2: Connectedness ⇔ flow feasibility. We claim that

$$\mathcal{H}_G(c) \text{ is connected} \iff \begin{cases} \forall v \in V_B^+, \; (\mathrm{FF}_v^+) \text{ is feasible,} \\ \forall v \in V_B^-, \; (\mathrm{FF}_v^-) \text{ is feasible.} \end{cases}$$

$\Longrightarrow$) Assume $\mathcal{H}_G(c)$ is connected. By Proposition 4.5, this is equivalent to

$$c \in \mathrm{Co}(\tilde{B}_{-i}) \qquad \forall i,$$

i.e. after removing any edge of the DAG, the balance equations remain solvable. For an out-bottleneck v, let us remove its unique outgoing edge e and take one non-negative solution $x \geq 0$ to the balance equation $\tilde{B}_{-e}x = c$, with $\tilde{B}_{-e}$ being the incidence matrix of G with the edge e removed.

Let us now consider the flow feasibility problem (FF_v^+) and show that it is feasible. All the internal and incoming arcs correspond to edges in G, meaning that we can associate to them the flows given by x. We then associate the flow $-c_u \geq 0$ to all the source arcs (s, u) and $c_u \geq 0$ to all the sink arcs (u, t), automatically satisfying the constraints of Equation (4.5). Finally, the

circulation arc carries a flow equal to the sum of the flows of all sink arcs $f_{(t,s)} = \sum_{u:c_u>0} c_u$. It is now possible to check that at every node $u \in N_v$ in the induced network, the flow conservation condition is achieved. In particular, at every hidden node u we have that the flows resulting from x on the internal arc give a balance value of c_u, which is then routed to either s or t depending on its sign, ensuring perfect balance. Thus, (FF_v^+) is feasible. The same reasoning can be applied to in-bottlenecks.

$\Longleftarrow$) Conversely, assume all $(\mathrm{FF}_v^\pm)$ are feasible, let e be any edge of the DAG and let us prove that $c \in \mathrm{Co}(\tilde{B}_{-e})$, i.e. there exists a nonnegative solution $x \geq 0$ to $\tilde{B}_{-e}x = c$.

If e is not the only outgoing/incoming edge of an out-/in-bottleneck, removing it still results in a network where all neurons belong to a path from the inputs V_I to the outputs V_O. This allows us to apply the non-emptiness result Proposition 4.4, resulting in $\tilde{B}_{-e}x = c$ having a non-negative solution.

Let e be the unique outgoing edge of an out-bottleneck v and let us explicitly build a non-negative solution $x \geq 0$ to $\tilde{B}_{-e}x = c$.

The problem (FF_v^+) is feasible, meaning that there exists a feasible flow f. For every edge e' in G_{-e} that belongs to the arc set A_v, we assign $x_{e'} = f_e$. These edges, we recall, are the "internal" ones connecting pairs of nodes inside $\overline{\mathrm{Anc}}(v)$ and the "incoming" ones, i.e. the edges connecting a node outside $\overline{\mathrm{Anc}}(v)$ to a pure ancestor of v. For all other edges, we initialize $x_{e'} = 0$. Observe now that, for every node $u \in \overline{\mathrm{Anc}}(v)$ the balance equation $(\tilde{B}_{-e}x)_u = c_u$ is satisfied due to the constraints in (FF_v^+) being satisfied (Equation (4.5)). In all other nodes which are not pure ancestors, the balance equations do not need to be satisfied.

Therefore, we apply the same reasoning of the proof of Proposition 4.4 and edit x to make $\tilde{B}_{-e}x = c$. To do this, we consider the fact that, for every node $u \in \tilde{V}$ that is not a pure ancestor of v, there exists a path p_I^u from inputs V_I to u and a path p_O^u from u to the outputs V_O.

If we denote by $\mathbb{1}_p$ the indicator function assigning 1 to all the edges belonging to path p and 0 to all others, we can see that $x + \alpha\mathbb{1}_{p_I^u} + \beta\mathbb{1}_{p_O^u}$ acts on the balance values in the following way:

- In all nodes $w \neq u$ the balance is left unchanged because either w does not belong to the edited paths and thus its balance is left untouched, or, if it does, the edited path adds the same quantity α or β to both the incoming and outgoing edges of w. Thus, in both cases $(\tilde{B}_{-e}(x + \alpha\mathbb{1}_{p_I^u} + \beta\mathbb{1}_{p_O^u}))_w = (\tilde{B}_{-e}x)_w$.

- In u, the edited paths add a quantity α to its incoming edges and a quantity β to its outgoing edges, thus $(\tilde{B}_{-e}(x + \alpha \mathbb{1}_{p_I^u} + \beta \mathbb{1}_{p_O^u}))_u = \alpha - \beta + (\tilde{B}_{-e}x)_u$.

This means that, by tuning $\alpha, \beta \geq 0$, we are able to satisfy the balance equation at u $(\tilde{B}_{-e}(x + \alpha \mathbb{1}_{p_I^u} + \beta \mathbb{1}_{p_O^u}))_u = c_u$ without altering the balance at other nodes. If we repeat this process for all other nodes $u \notin \overline{\mathrm{Anc}}(v)$, we find a non-negative solution

$$\hat{x} = x + \sum_{u \notin \overline{\mathrm{Anc}}(v)} \alpha_u \mathbb{1}_{p_I^u} + \beta_u \mathbb{1}_{p_O^u} \geq 0$$

that satisfies $\tilde{B}_{-e}\hat{x} = c$. Hence $c \in \mathrm{Co}(\tilde{B}_{-e})$ for all e, and connectedness follows from Proposition 4.5.

Step 3: Feasibility via cut conditions. We now conclude the proof by deriving a more interpretable condition that is equivalent to the network flow feasibility problem $(\mathrm{FF}_v^{\pm})$. To do this, we resort to the following classical result in network flow theory.

Proposition 4.7 (Hoffman's theorem [292], reported in Ref. [291])**.** *A network (N, A) with conservation and boundedness constraints $l_{ij} \leq f_{ij} \leq m_{ij}$ with non-negative lower bounds $l_{ij} \geq 0$ is feasible if and only if for every non-trivial partition (S, T) $S \sqcup U = N$, called a* cut*, we have $V(S, T) \leq 0$, where:*

$$V(S, T) = \sum_{\substack{i \in S \\ j \in T}} l_{ij} - \sum_{\substack{i \in T \\ j \in S}} m_{ij}. \tag{4.6}$$

We recall that a partition (S, T) is trivial when either $S = \emptyset$ or $T = \emptyset$. Fix a bottleneck v and consider the associated flow feasibility problem (FF_v) (with node set N and arc set A). Recall that (FF_v) is a circulation problem with lower and upper bounds l_{ij}, m_{ij} on each arc $(i, j) \in A$, and that all lower bounds are nonnegative (Equation (4.5)). We now specialize the condition of Proposition 4.7 to the structure of (FF_v). In particular, we seek for conditions for (S, T) to be a positive cut, i.e. guaranteeing the non feasibility of (FF_v).

First, observe that the circulation arc (t, s) has infinite capacity. Hence, if $s \in S$ and $t \in T$ the second term in (4.6) equals $-\infty$, so $V(S, T) \leq 0$ automatically. Moreover, if $s \in T$ and $t \in S$, then the cut cannot be positive, because all the arcs with strictly positive lower bounds (i.e. the source and sink arcs) either come out of s or go into t. Therefore, any cut with $V(S, T) > 0$ must place s and t on the same side of the partition.

Next, note that all internal and incoming arcs between nodes of $N \setminus \{s\} = \overline{\mathrm{Anc}}(v) \cup \{s\}$ have infinite upper bound. Thus, if a cut (S, T) contains an internal or incoming arc directed from T to S, the second sum in (4.6) is again infinite, and $V(S, T) \leq 0$. Consequently, any cut with $V(S, T) > 0$ must satisfy the following closure property:

$$\text{if } u \in T \text{ and } (u, w) \in A \text{ is an internal or incoming arc then } w \in T.$$

Equivalently, $T \cap (\overline{\mathrm{Anc}}(v) \cup \{s\})$ must be *closed by forward arcs*.

Moreover, this means that $\{s, t\} \subseteq S$ because, if $s \in T$ then, by closedness by forward arcs, this would imply that T would contain all nodes and (S, T) would be a trivial cut.

These two observations imply that any cut with positive value is necessarily of the form

$$S = N \setminus T, \qquad T \subset \overline{\mathrm{Anc}}(v) \text{ forward closed.}$$

Let $T \subset \overline{\mathrm{Anc}}(v)$ be forward stable and set $S = N \setminus T$. We compute $V(S, T)$.

The only arcs contributing to the first sum in (4.6) are source arcs (s, u) with $u \in T$ and $c_u < 0$, each having lower bound $-c_u$. Hence

$$\sum_{\substack{i \in S \\ j \in T}} l_{ij} = \sum_{\substack{u \in T \\ c_u < 0}} (-c_u).$$

By forward closedness of T, the only arcs from T to S are sink arcs (u, t) with $u \in T$ and $c_u > 0$, each having upper bound c_u. Therefore

$$\sum_{\substack{i \in T \\ j \in S}} m_{ij} = \sum_{\substack{u \in T \\ c_u > 0}} c_u.$$

Combining the two expressions in Equation (4.6) yields

$$V(S, T) = \sum_{\substack{u \in T \\ c_u < 0}} (-c_u) - \sum_{\substack{u \in T \\ c_u > 0}} c_u = -\sum_{u \in T} c_u.$$

Conclusion. We conclude that (FF_v^+) is infeasible if and only if there exists a forward-cloesd subset $T \subset \overline{\mathrm{Anc}}(v)$ such that

$$\sum_{u \in T} c_u < 0.$$

Equivalently,

$$(\mathrm{FF}_v^+) \text{ is feasible } \iff \forall T \subset \overline{\mathrm{Anc}}(v) \text{ forward closed}, \quad \sum_{u \in T} c_u \geq 0.$$

The corresponding statement for (FF_v^-) follows by reversing all arrows and replacing c with $-c$.

□

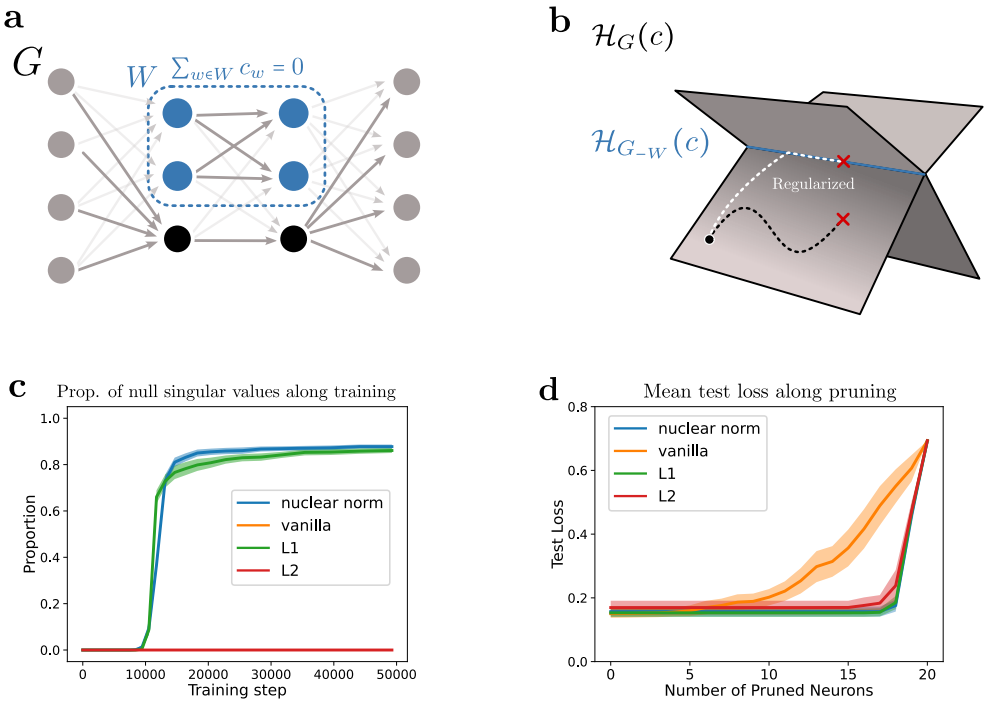

Fig. 4.4 **a.** A singularity of the invariant set corresponds to a configuration where a set of neurons is cut out from input and outputs. **b.** Visualization of the training dynamics on an invariant set with singularities. **c.** Proportion of null singular values along training for shallow network with 20 hidden neurons with and without regularization. **d.** Test losses as a function of the number of neurons pruned. Shaded regions in **c.** and **d.** denote confidence intervals over 50 independent trainings that have converged to a low loss solution.

4.3.3 Singularities

While singularities in the neuromanifold—i.e., the function space—have been extensively studied [279, 254, 280, 293, 294], in this section we propose a complementary perspective by analyzing singularities of the invariant set $\mathcal{H}_G(c)$, which sits in the optimization (parameter) space.

Singularities are subnetworks. In algebraic geometry, a *singularity* refers to a "degenerate" point in an algebraic variety where the tangent space has a dimension greater than the dimension of the variety. If a variety is given by the system of n polynomial equations $g(x) = 0 \in \mathbb{R}^n$, singularities correspond to the points x where the Jacobian matrix $J(x) = \nabla_x g(x)$ has its rank reduced from the maximal possible. When the rank of the Jacobian is maximal, the point is instead said to be *regular*.

In the case of the invariant set, the Jacobian matrix is computed by taking the derivative of $\tilde{B}\theta^2 - c$ w.r.t. θ, resulting in

$$J_G(\theta) = 2\tilde{B}\,\mathrm{diag}(\theta). \tag{4.7}$$

From Equation (4.7), we see how the Jacobian matrix has the same structure of the graph incidence matrix $\tilde{B}$, with each of the edges (the columns) weighted by its associated parameter θ. Therefore, it follows that the only way in which the matrix can have a lower rank is when some of the values of θ are 0, i.e., when some of the edges are effectively cut from the computational graph.

In the theorem below, we completely characterize the singularities of the invariant set in terms of the presence of sets of parameters equal to zero that fundamentally change the topology of G.

Theorem 4.2 (Singularities disconnect neurons)**.** *If $G/\partial V$ is the undirected (multi-) graph obtained from G by gluing together input and output nodes and neglecting edge direction, and $\mathcal{E}(\theta) \subseteq E$ is the set of edges with zero weight ($\theta_e = 0 \iff e \in \mathcal{E}$), then*

$$\mathrm{rank}\, J_G(\theta) = |\tilde{V}| + 1 - \beta_0\left(G_{-\mathcal{E}(\theta)}/\partial V\right),$$

where $\beta_0(G)$ is the number of connected components of G.

Proof. Let us compute $\mathrm{rank}\, J_G(\theta)$. First, observe that if $\theta_e \neq 0\ \forall e \in E$, then $\mathrm{rank}\, J_G(\theta) = \mathrm{rank}(\tilde{B})$ as $\mathrm{diag}(\theta)$ is invertible. Therefore, the rank of the Jacobian can decrease if and only if some parameters are 0, i.e. when some edges are effectively removed from the computational graph.

Define now $\mathrm{diag}(\theta)^{\dagger}_{ee} = 1/\theta_e$ if $\theta_e \neq 0$ and 0 otherwise, and $\mathcal{E}(\theta) = \{e \in E | \theta_e = 0\}$ the set of zero-weight edges. Observe that

$$\mathrm{rank}\, J(\theta) = \mathrm{rank}(\tilde{B}\mathrm{diag}(\theta)) = \mathrm{rank}(\tilde{B}\,\mathrm{diag}(\theta)\,\mathrm{diag}(\theta)^{\dagger}) = \mathrm{rank}(\tilde{B}_{-\mathcal{E}(\theta)}),$$

where $\tilde{B}_{-\mathcal{E}}$ is $\tilde{B}$ with the columns corresponding to edges in $\mathcal{E}$ put to 0. Now, it holds that

$$\mathrm{rank}(\tilde{B}_{-\mathcal{E}(\theta)}) = \mathrm{rank}(\tilde{B}^{\top}_{-\mathcal{E}(\theta)}) = |\tilde{V}| - \dim\ker(\tilde{B}^{\top}_{-\mathcal{E}(\theta)}), \tag{4.8}$$

where the last equality follows from the rank nullity theorem.

To relate $\dim\ker(\tilde{B}^\top_{-\mathcal{E}(\theta)})$ to the topological properties of G, we employ the interpretation of $\tilde{B}$ as the matrix representation of the relative boundary operator of the pair $(G, \partial V)$, as outlined in Example 0.4 and Remark 4.3.

Therefore, $\ker \tilde{B}^\top$ can be interpreted as the *0-th relative co-homology group* $H^0(G, \partial V)$ of the pair $(G, \partial V)$. Given that we are considering real coefficients, we can equivalently employ the homology group $H^0(G, \partial V) \cong H_0(G, \partial V)$ due to Proposition 0.4, meaning that in both cases we are measuring the number of relative connected components of the space. We can finally resort to Proposition 0.2 and rewrite the relative homology as the homology of the quotient space

$$\dim H^0(G_{-\mathcal{E}(\theta)}, \partial V) = \dim H_0(G_{-\mathcal{E}(\theta)}, \partial V) = \dim \tilde{H}_0(G_{-\mathcal{E}(\theta)}/\partial V) = \beta_0(G_{-\mathcal{E}(\theta)}/\partial V) - 1,$$

where, recall, $\tilde{H}_0$ is the *reduced* homology group (Equation (5)), $G/\partial G$ is the graph G with nodes in ∂V all glued together, and $G_{-\mathcal{E}(\theta)}$ is the graph G with the edges in $\mathcal{E}(\theta)$ removed. We can therefore go back to Equation (4.8) and state that

$$\operatorname{rank} J(\theta) = |\tilde{V}| - \dim \tilde{H}_0(G_{-\mathcal{E}(\theta)}/\partial V) = |\tilde{V}| - \beta_0(G_{-\mathcal{E}(\theta)}/\partial V) + 1,$$

meaning that the rank of the Jacobian is less than its maximum value of $|\tilde{V}|$ only when the number of connected components of the quotient graph $G_{-\mathcal{E}(\theta)}/\partial V$ is greater than 1. This happens if and only if removing edges in $\mathcal{E}(\theta)$ disconnects a set of nodes from both input and output neurons. □

Therefore, it follows from Theorem 4.2 that $\operatorname{rank} J_G(\theta) = |\tilde{V}|$ for regular parameters, and its rank decreases for parameters whose zero edges disconnect some hidden neurons from both input and output neurons. If a group of neurons is such that it is not connected by any path to both input and output neurons, it means that it is a useless component of the network, as it takes no part in any computation, both in the feed-forward and in the back-propagation phases. Singularities, therefore, correspond to effective *subnetworks* of the original neural network, as shown on Figure 4.4 a. This echoes similar results connecting subnetworks to the singular points of the neuromanifold [295–297, 294].

This observation can be leveraged to prove the following result giving us the dynamical behavior of singularities of the invariant set.

Proposition 4.8 (Singularities are invariant under GF)**.** *Let the neural network be trained with gradient flow, $\dot{\theta} = -g(\theta)$, $\theta(0) \in \mathcal{H}_G(c)$ and $W \subseteq \tilde{V}$ be a set of hidden neurons, with $\mathcal{E} \subseteq E$*

the set of edges that connecting neurons in W to neurons outside of W and vice versa. If for a time t_0, $\theta_e(t_0) = 0$ for every $e \in \mathcal{E}$ (meaning that θ is singular in $\mathcal{H}_G(c)$), then, for any $t > t_0$, $\theta_e(t) = 0$ for every $e \in \mathcal{E}$.

Proof. To prove the result, we will prove that, if $\theta_e(t_0) = 0$ for every edge $e \in \mathcal{E}$, the same holds for the gradients: $g_e(t_0) = 0$ for every edge $e \in \mathcal{E}$, meaning that gradient flow will not update their parameters, leaving them fixed at 0.

Let us denote by $\mathcal{P}_{v,V_O}$ the set of all paths from node v to a node in V_O i.e. $p \in \mathcal{P}_{v,V_O} \iff p = (u_1, \dots, u_{n_p})$ with $u_1 = v$, $u_{n_p} \in V_O$ and $(u_i, u_{i+1}) \in E$, for all i.

Let a_v, z_v be the activation and pre-activation of neuron v, respectively

$$z_v = \sum_{(u,v)\in E} \theta_{(u,v)} a_u, \quad a_v = \sigma(z_v).$$

The chain rule allows us to decompose the gradient of the loss flowing through neuron v in the contributions of all path from v to an output neuron as follows:

$$\frac{\partial L}{\partial a_v} = \sum_{p=(u_1,\dots,u_{n_p})\in\mathcal{P}_{v,V_O}} \frac{\partial L}{\partial a_{u_{np}}} \prod_{i=2}^{n_p} \frac{\partial a_{u_i}}{\partial z_{u_i}} \frac{\partial z_{u_i}}{\partial a_{u_{i-1}}} \tag{4.9}$$

For any neuron in the disconnected set $v \in W$, it holds that any path to the output $p \in \mathcal{P}_{v,V_O}$ contains an edge e_p such that $\theta_{e_p} = 0$. If we notice that $\frac{\partial z_{u_i}}{\partial a_{u_{i-1}}} = \theta_{(u_{i-1},u_i)}$, we have that, for any path p, the product in Equation (4.9) will contain e_p and therefore its value will be 0 $\frac{\partial L}{\partial a_v} = 0 \; \forall v \in W$.

Let us now prove that this implies that the edges that disconnect W have zero gradient.

First, let $e = (u, w) \in E$ with $u \in V \setminus W$ and $w \in W$.

$$g_e := \frac{\partial L}{\partial \theta_e} = \frac{\partial L}{\partial a_w} \frac{\partial a_w}{\partial \theta_e} = 0,$$

because $w \in W$.

Let $e = (w, v) \in E$ with $w \in W$ and $v \in V \setminus W$.

$$g_e = \frac{\partial L}{\partial \theta_e} = \frac{\partial L}{\partial a_v} \frac{\partial a_v}{\partial \theta_e} = \frac{\partial L}{\partial a_v} \frac{\partial a_v}{\partial z_v} \frac{\partial z_v}{\partial \theta_e},$$

where $\frac{\partial z_v}{\partial \theta_e} = a_w$. Given that W is also disconnected from the input nodes V_I, any node inside must have 0 activations. Therefore $w \in W \implies a_w = 0$ and $g_e = 0$. □

Proposition 4.8 means that if the parameter reaches a singularity, then it cannot escape from it: once a network module has been killed, it cannot be revived.

Singularities are rare. One would be tempted to think that the presence and invariance of singularities could provide the explanation for the neural network performing an automatic model selection through the progressive movement from one singularity to another, smaller one. We show here, however, that this picture doesn't hold for the singularities of the invariant set for two reasons: **1.** Given a random initialization, the probability of $\mathcal{H}_G(c)$ having singularities is 0. **2.** If $\mathcal{H}_G(c)$ has singularities, then the learning trajectory cannot reach them in finite time.

Proposition 4.9. *Let $c \in \mathbb{R}^{|\tilde{V}|}$. If $\mathcal{H}_G(c)$ admits singularities then there exists a subset of hidden neurons $W \subseteq \tilde{V}$ such that $\sum_{v\in W} c_v = 0$.*

Proof. By definition, a singularity identifies a disconnected set of nodes $W \subseteq \tilde{V}$. If we denote the edges inside W by $E_W = \{e = (u, v) \in E \mid u, v \in W\}$, we have that $\sum_{v\in W} c_v = \sum_{e\in E_W} \theta_e^2 - \theta_e^2 = 0$. This is because each edge in E_W is shared by exactly 2 nodes of W and all other edges in or out of W have weight 0. □

To have a singularity, therefore, an exact equality condition on c must hold. (Figure 4.4). If we sample the initial parameter with any initialization scheme where each parameter is independently sampled from the real numbers $\mathbb{R}$, we see that the probability that a set of neurons will have sum *exactly* zero must be zero. Moreover, a stronger statement than the one of Proposition 4.8 can be derived.

Proposition 4.10. *Under GF optimization, we have that* $\operatorname{rank} J(\theta(0)) = \operatorname{rank} J(\theta(t)) \; \forall\, 0 \leq t < \infty$.

Proof. Let $\dot{\theta}(t) = -g(\theta(t))$ be the evolution of the parameter configuration under gradient flow. Let $J_G(\theta) = 2\tilde{B}\,\mathrm{diag}(\theta)$ be the Jacobian matrix of the invariant set $\mathcal{H}_G(c)$ and let us derive the evolution equation for J_G.

$$\frac{\mathrm{d}}{\mathrm{d}t} J_G(t) = \frac{\mathrm{d}}{\mathrm{d}t} 2\tilde{B}\,\mathrm{diag}(\theta(t)) = 2\tilde{B}\,\mathrm{diag}(\dot{\theta}(t)) = -2\tilde{B}\,\mathrm{diag}(g(\theta(t))). \tag{4.10}$$

It turns out that we can simplify this equation by leveraging the connections between rescaling symmetries and GF dynamics. In fact, we can take the gradient balance equation of Proposition 4.2 $\tilde{B}(\theta \odot g(\theta)) = 0$ and differentiate it w.r.t. θ to obtain

$$\tilde{B}\,\mathrm{diag}(\theta)H(\theta) + \tilde{B}\,\mathrm{diag}(g(\theta)) = 0. \tag{4.11}$$

Plugging Equation (4.11) in Equation (4.10), we finally get

$$\frac{\mathrm{d}}{\mathrm{d}t}J_G(t) = 2\tilde{B}\,\mathrm{diag}(\theta)H(\theta) = J_G(\theta)H(\theta), \tag{4.12}$$

i.e. the evolution of the Jacobian matrix is dictated by the Hessian matrix of the loss function through right matrix multiplication.

It is known (e.g. [298] page 15) that the solution of Equation (4.12) can be written in the following way

$$J_G(t) = J_G(0)\mathcal{T}\exp(\int_0^t H(s)\mathrm{d}s),$$

where $\mathcal{T}\exp(\int_0^t H(s)\mathrm{d}s)$ is the time-ordered matrix exponential defined as

$$\mathcal{T}\exp(\int_0^t H(s)\mathrm{d}s) = \sum_{n=0}^{\infty}\int_0^t \mathrm{d}t_1' \int_0^{t_1'} \mathrm{d}t_2' \cdots \int_0^{t_{n-1}'} \mathrm{d}t_n' H(t_1')\cdots H(t_n')$$

which ensures that the terms of the exponential series are in the right order, as the matrices $H(t_1)$ $H(t_2)$ may not commute.

Finally, in Theorem 4 of Ref. [298], it is shown that the time-ordered matrix exponential can be written as a standard matrix exponential and is therefore invertible for any finite t. This means that $\mathrm{rank}(J_G(t)) = \mathrm{rank}(J_G(0))$, thus concluding the proof.

□

This result of Proposition 4.9 tells us that a gradient flow trajectory cannot fall into a singularity in finite time. Together, Propositions 4.9 and 4.10 implies that singularities generally don't exist in the optimization space when using common initializations and, even with a specifically chosen initialization ($c = 0$), they are effectively unreachable under gradient flow. Note, however, that this does not exclude that other learning schemes, such as finite-time gradient descent, may be able to reach them.

Inducing singularities. As shown above, if the training were able to reach singularities, it would allow the model to perform a "self-pruning" process, where useless network modules are automatically killed and removed from the computation. To actively drive the training dynamics towards them, we explore the use of a singularity-promoting regularization, enabling us to enforce differentiable pruning in a very general way, entirely agnostic to the DAG topology. The idea is illustrated by Figure 4.4 b. A natural approach to target singularities directly from their definition is to penalize the rank of the Jacobian of $\mathcal{H}_G(c)$, $J_G(\theta) = 2\tilde{B}\,\mathrm{diag}\,(\theta)$. However, since the rank is a non-differentiable function, we instead penalize the *nuclear norm* of the Jacobian $\|J_G(\theta)\|_*$—the sum of its singular values— which is known to be a smooth surrogate of the rank function [299].

Numerical experiments As an illustrative example, we test our approach on the well-known Breast Cancer classification dataset [300] using a range of MLP architectures—shallow and deep, with or without biases and skip connections. To approximate continuous gradient flow, we train with stochastic gradient descent (SGD) using a small step size (0.001), comparing nuclear norm regularization against L1 and L2. Tracking the Jacobian rank during training confirms that the nuclear norm consistently drives the model toward singularities (Figure 4.4c). For instance, a shallow network can disconnects around 18 of its 20 hidden units, whereas L2 and unregularized training leave all neurons active. Surprisingly, L1 regularization performs similarly to the nuclear norm, despite being traditionally associated with parameter (and not neuron) sparsity. This suggests that L1 may implicitly promote singularities and, while a precise theoretical understanding is outside the scope of this chapter, we provide an empirical analysis in Section 4.3.5. Finally, reaching singular configurations guarantees the ability to perform lossless pruning, and while L2 is already quite robust to pruning, disconnecting active neurons introduces modifications to the implemented function (Figure 4.4d).

4.3.4 Experimental details

Here we provide additional details regarding experiments described in the main paper.

Disconnected neurons and pruning

Figures 4.4c–d show, respectively, the proportion of null singular values during training and the evolution of test loss under neuron pruning for a shallow, bias-free architecture. We report in Figure 4.5 results for a broader range of architectures discussed in the main text. Both nuclear

norm and $L1$ regularization promoted neuron sparsity (fig. 4.5, left column). Note that skip connections made neuron pruning more challenging, as illustrated in previous literature [276]. Below, we describe our two pruning strategies; however in both cases, nuclear norm and $L1$ remain the most robust, vanilla the most fragile, and $L2$ intermediate (fig. 4.5, center and right columns).

Architectural details. We evaluate six architectures: a shallow network with 20 hidden units, with and without bias; and four multilayer perceptrons (MLPs) with three layers of 10 hidden units, varying by the presence of bias and whether a skip connection links the first and second hidden layers.

Details on methodology. All architectures were trained from scratch for 5000 epochs using SGD with a learning rate of 10^{-3}. Regularization strengths were empirically tuned to balance low task loss and singular value minimization: $\alpha_{nuc} = 0.05$ (nuclear norm), $\alpha_{L1} = 10$, and $\alpha_{L2} = 20$. The total loss is defined as $L = L_{task} + \alpha_{reg} L_{reg}$. For each architecture, we sampled 30 runs that converged to low train loss models to analyze training dynamics and pruning robustness. Singular values below 10^{-3} were considered null. The dataset was not preprocessed. All experiments ran on CPU over approximately 50 hours.

Pruning. We iteratively remove entire neurons (i.e., groups of parameters) from trained models and measure the degradation in performance, as measured by the test loss. For each hidden neuron v, we compute a principled pruning score

$$s_v = \left(\sum_{u:(u,v)\in E} \theta^2_{(u,v)} \right) \left(\sum_{w:(v,w)\in E} \theta^2_{(v,w)} \right) \tag{4.13}$$

which is the product of the L2 norms of its input and output weights. This score is low for (nearly) disconnected neurons and invariant under neuron rescaling, making it robust to reparameterization. However, a max-based score,

$$s'_v = \max \left\{ \sum_{u:(u,v)\in E} \theta^2_{(u,v)}, \sum_{w:(v,w)\in E} \theta^2_{(v,w)} \right\} \tag{4.14}$$

produced slightly different results and is reported as well for comparison.

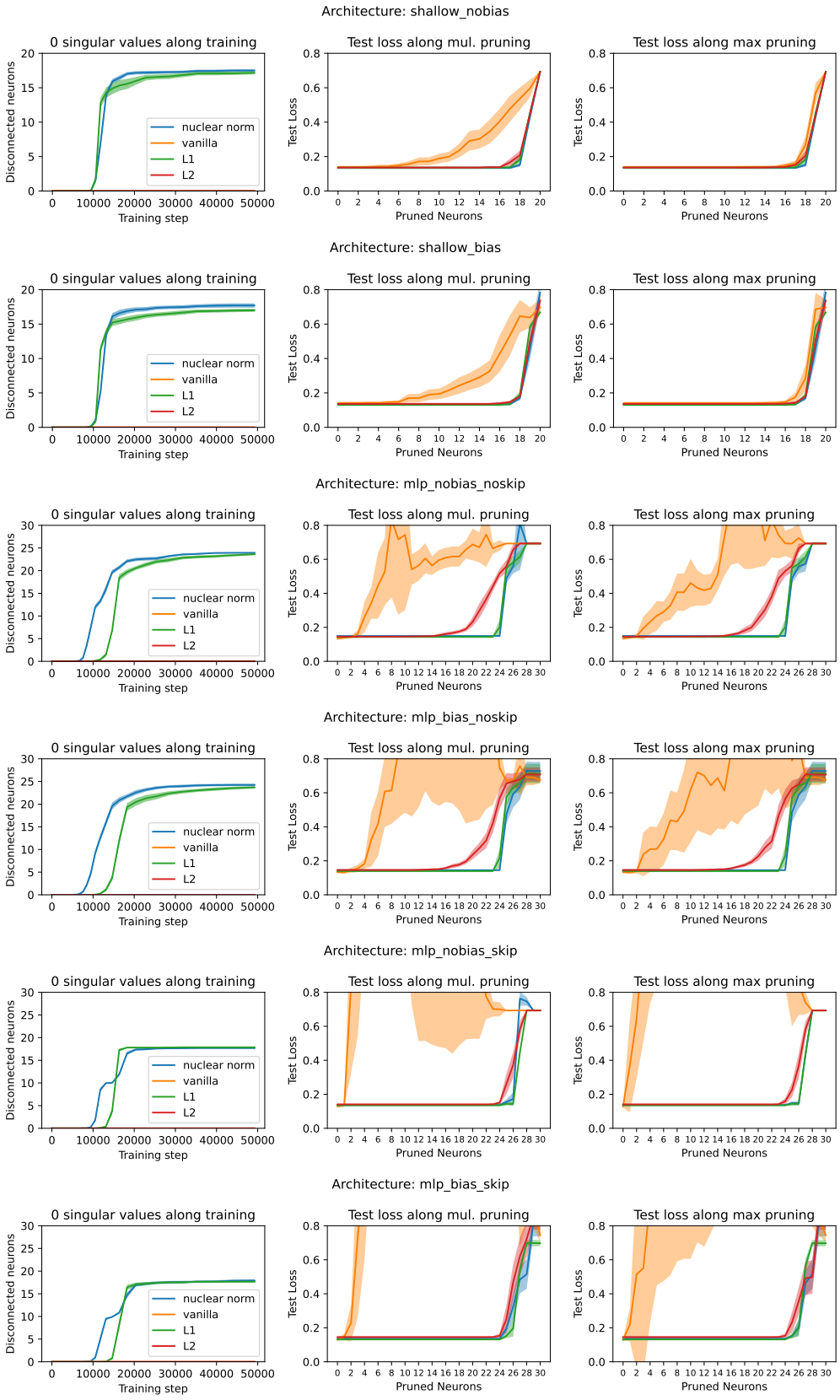

Fig. 4.5 **(Rows)** architectures **(Left column)** number of almost 0 singular values (threshold: 10^{-3}) along training. **(Center column)** pruning neurons on trained networks using s_v (multiplicative) and **(Right column)** s'_v (maximum) scores

Training dynamics

For one training of shallow bias-free network, we present the evolution of key quantities during training to highlight how different regularizations affect their behavior.

As shown in Figure 4.6, both nuclear norm and L_1 regularization drive many neurons' balance numbers c_k to zero—consistent with the fact that disconnecting a neuron in a fully connected layer requires its input and output weights to vanish, making c_k null by definition (Equation (3.14)). Without regularization, c_k values, which should be fixed under continuous gradient flow, can increase due to the effect of discrete optimization steps. These steps cause a gradual drift out of $\mathcal{H}_G(c)$, as the update vector deviates from the ideal optimization path. Notably, in this unregularized case, c_k values remain more stable. Figure 4.7 present the complementary view of singular values.

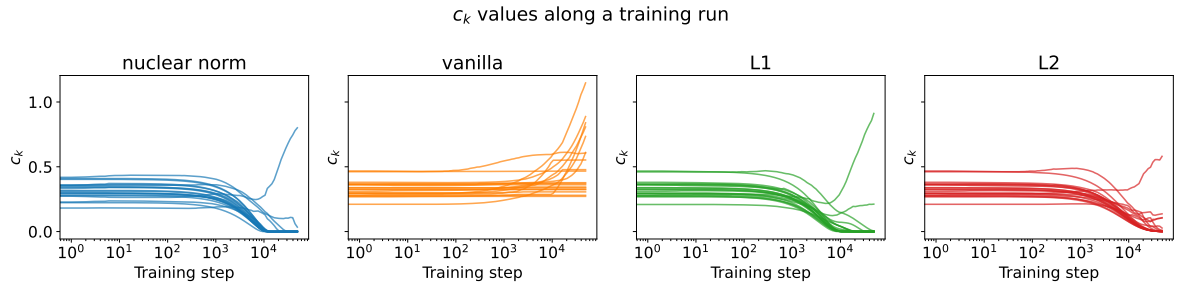

Fig. 4.6 Balance values c_k during training of a shallow, bias-free network. $k \in [0, 20]$ is the neuron index.

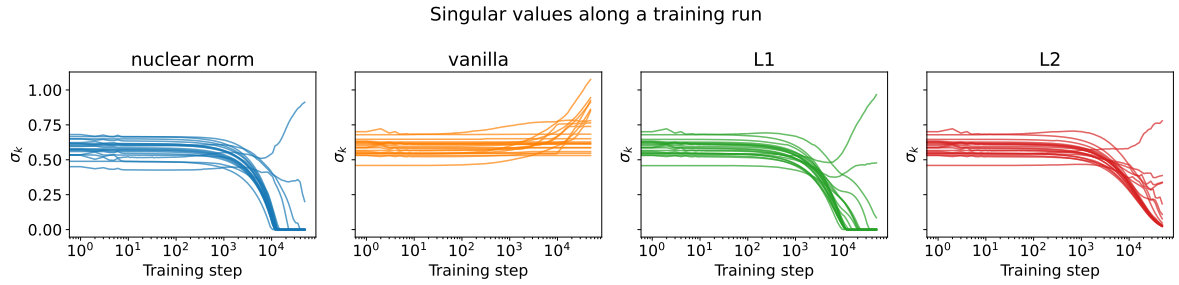

Fig. 4.7 Singular values of the Jacobian $\tilde{B}\,\mathrm{diag}(\theta)$ during training of a shallow, bias-free network. $k \in [0, 20]$ is the neuron index.

4.3.5 Empirical comparison between L1 and nuclear regularization

In the main text we observe that both the nuclear norm and L1 regularizers prune a similar number of neurons (see also Figure 4.4c, Figure 4.5). Here, we give empirical evidence on key distinctions between these regularizers. We first show that both regularizers cannot be explained by a null model and then illustrate that the nuclear norm preserves edges on active neurons, achieving strong group regularization while L1 is more aggressive, also pruning edges belonging to active neurons.

Null model

In this part we use a simple null model to rule out the explanation that L1 regularization achieves neuron sparsity solely due to its known edge sparsity mechanism. The null model works as follow: we estimate the probability p_{L1} for a generic edge to be dropped after training. Intuitively, we choose an edge before training and observe after training under L1 regularization if it was dropped or not. To decide if the edge was dropped or not, we use a threshold of 10^{-3} which corresponds to a clear peak in the parameters values distribution, stable over multiple orders of magnitude. Then, starting from the initial computational graph (i.e. before training, with all edges), we can compute analytically the expected number of disconnected neurons if every edge is dropped with probability p_{L1}. Results are reported on Figure 4.8 left. We repeat the same analysis for the nuclear norm regularization, using the same threshold we obtain another null model with probability of dropping a random edge of p_{nuc} and report the expected number of disconnected neurons on Figure 4.8 right. In summary, both null models cannot explain the number of pruned neurons by the number of pruned edges, meaning that in both cases there must be other underlying mechanisms. The underlying mechanism is explicit in the case of the nuclear norm regularization, since neurons are directly targeted, but remains veiled in the case of L1. Note also that the null model is an even worse explanation in the singular regularization case, indicating a stronger alternative mechanism.

Active weights and pruned neurons

To further investigate the difference between L1 and nuclear norm, we turn our attention to the distribution of parameters magnitude, which is plotted on Figure 4.9 left for 4 representative trainings. For the singular regularization we observe a clear separation in the magnitude of parameters belonging to pruned and active neurons, and this is not the case for L1 regularization.

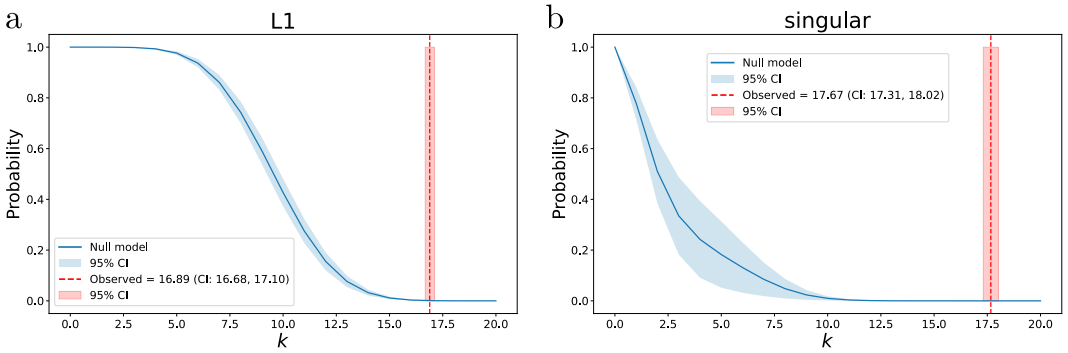

Fig. 4.8 Probability of having at least k pruned neurons under the null model for **(a)** nuclear norm regularizer **(b)** L1 regularizer. The red dotted vertical line is the observed number of pruned neurons, the blue curve is the analytic probability of pruning at least k neurons under the null model.

This means that the nuclear norm either drops completely a neuron (i.e. all its edges at the same time) or keeps the neuron active. In contrast, the separation is less clear for L1: there are many inactive weights on active neurons. Figure 4.9 right shows the aggregated results for 30 models of each type, where we observe an overlap of the distribution for L1 only.

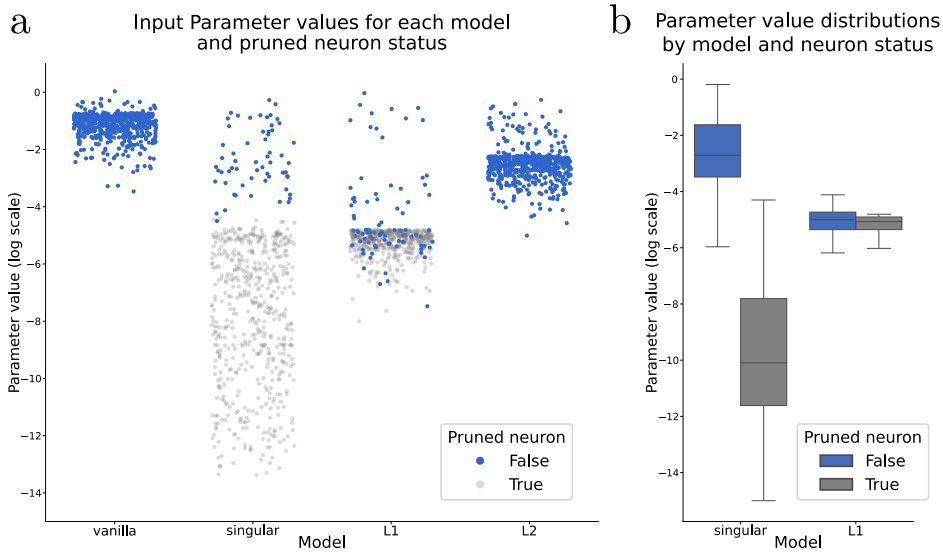

Fig. 4.9 **a.** Absolute value of input parameters for the hidden layer of a shallow network with bias. **b.** Absolute value of parameters (including bias) for the same network architecture on 30 independent runs.

4.4 Conclusion

In this chapter, we extended the results of Chapter 3, exploring two underexplored pathologies in the training of ReLU neural networks defined over general DAGs: the (dis)connectedness of the invariant parameter space and the emergence of singularities within it. By leveraging the neuron-wise rescaling invariance of homogeneous activations and analyzing the associated conservation laws under gradient flow, we provided a complete characterization of the invariant set as an algebraic variety constrained by quadratic balance conditions.

Our topological analysis revealed that disconnections in the optimization space are dictated by the presence of bottleneck neurons and an imbalance in flow conditions. We further demonstrated that singularities correspond to effective subnetworks, and although gradient flow trajectories cannot reach them in finite time, their role can be leveraged to improve structured pruning. We introduced a nuclear norm regularizer that promotes convergence toward such configurations. Surprisingly, we observed that L1 regularization can induce comparable effects, hinting at a deeper connection between sparsity and singularity-driven pruning.

Back to Table of Contents

Chapter 5

Conclusions

This thesis has investigated the role of topology and geometry in the analysis of complex systems, with a focus on two settings that are a priori distant: higher-order dynamical systems on networks and the training dynamics of ReLU neural networks. Despite the differences in motivation and modeling assumptions, the results developed in the two parts of the thesis are connected by a common mathematical viewpoint, centered on the use of algebraic topology to characterize collective behavior and constraints arising from the underlying space.

The first part of the thesis addressed dynamical processes on higher-order networks, modeled through simplicial complexes. Starting from existing formulations of simplicial Kuramoto-type models, the contribution of this work was not to introduce a new class of models, but rather to provide a unified and systematic analytical framework for their study. By expressing these dynamics in terms of operators acting on simplicial cochains, it was possible to give a clear interpretation of the different interaction mechanisms that arise across simplicial orders, distinguishing the roles of lower- and upper-adjacency couplings. This perspective made it possible to characterize the equilibrium structure of the dynamics and to derive explicit stability conditions for phase-locked states. In particular, the analysis highlighted how the topology of the underlying simplicial complex constrains the set of admissible synchronized configurations and influences their robustness.

This operator-based viewpoint also provides the natural link between the first two chapters of the thesis. In both cases, the dynamics under consideration are governed by Laplacian-like operators acting on simplices, albeit with different interpretations. In the simplicial Kuramoto setting, interactions are mediated by a nonlinear extension of the Hodge Laplacian, whose kernel

contains the fully synchronized states. This same emphasis on Laplacian-driven processes motivates the second chapter, where diffusion is no longer confined to a fixed simplicial order but is allowed to occur across dimensions through the introduction of cross-order Laplacians. One of the outcomes of this comparison is the observation that standard Hodge-based diffusion, while mathematically natural, does not always correspond to a physically meaningful notion of diffusion on higher-order structures. The cross-order construction addresses this limitation by explicitly encoding interactions between simplices of different dimensions.

Within this framework, the second chapter developed a renormalization scheme for higher-order networks based on spectral and entropic observables associated with cross-order Laplacians. This approach extends classical ideas from statistical physics to simplicial complexes and provides a principled way to probe their multiscale organization. The resulting renormalization procedure was shown to capture non-trivial scale-invariant behavior in both synthetic models and empirical datasets, with different simplicial orders exhibiting distinct scaling signatures.

The second part of the thesis shifted focus to the topology and geometry of the parameter space of ReLU neural networks trained by gradient-based methods. Here, the starting point was the observation that positive homogeneity of the ReLU activation induces continuous rescaling symmetries, leading to conserved quantities along training trajectories. By characterizing the associated invariant sets as algebraic varieties, the thesis provided a detailed analysis of their topological properties. For shallow networks, it was shown that these invariant sets can be disconnected, giving rise to a genuine topological obstruction to training that depends on initialization. Exact conditions for disconnectedness were derived, together with quantitative estimates of the probability of encountering such obstructions under standard random initializations.

These results were then extended to general feedforward architectures modeled as directed acyclic graphs. In this setting, the invariant set was described as a higher-dimensional algebraic variety whose connectedness depends on architectural features such as bottlenecks and imbalances in the flow of weights. A complete characterization of connectedness was obtained, together with an explicit description of the singularities of this space. These singular points were shown to correspond to effective subnetworks, providing a precise geometric interpretation of structured pruning phenomena. The analysis further clarified how different regularization schemes interact with the geometry of the invariant set, with nuclear norm regularization promoting convergence toward low-rank, singular configurations.

References

[1] Nicholas W Landry, Maxime Lucas, Iacopo Iacopini, Giovanni Petri, Alice Schwarze, Alice Patania, and Leo Torres. Xgi: A python package for higher-order interaction networks. *Journal of Open Source Software*, 8(85):5162, 2023.

[2] Leo J Grady and Jonathan R Polimeni. *Discrete calculus: Applied analysis on graphs for computational science*. Springer Science & Business Media, 2010.

[3] Marco Nurisso. *Analytical characterization of the simplicial Kuramoto model*. PhD thesis, Politecnico di Torino, 2022.

[4] J.R. Munkres. *Topology*. Featured Titles for Topology. Prentice Hall, Incorporated, 2000.

[5] Aliaksei Sandryhaila and José MF Moura. Discrete signal processing on graphs. *IEEE transactions on signal processing*, 61(7):1644–1656, 2013.

[6] Allen Hatcher. *Algebraic topology*. Cambridge University Press, Cambridge, 2002.

[7] Arkady Kheyfets. The boundary of a boundary principle: A unified approach. *Foundations of physics*, 16(5):483–497, 1986.

[8] John M Lee. Smooth manifolds. In *Introduction to smooth manifolds*, pages 1–29. Springer, 2003.

[9] Nicholas Young. *An introduction to Hilbert space*. Cambridge university press, 1988.

[10] Beno Eckmann. Harmonische funktionen und randwertaufgaben in einem komplex. *Commentarii Mathematici Helvetici*, 17(1):240–255, 1944.

[11] Mark Newman. *Networks*. Oxford university press, 2018.

[12] Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small-world'networks. *nature*, 393(6684):440–442, 1998.

[13] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[14] Mauricio Barahona and Louis M Pecora. Synchronization in small-world systems. *Phys. Rev. Lett.*, 89(5):054101, 2002.

[15] A. Pikovsky, M. Rosenblum, and J. Kurths. *Synchronization: A Universal Concept in Nonlinear Sciences*. Cambridge University Press, 2003.

[16] S. H. Strogatz. *Sync: The Emerging Science of Spontaneous Order*. Penguin UK, 2004.

[17] Cameron Nowzari, Victor M Preciado, and George J Pappas. Analysis and control of epidemics: A survey of spreading processes on complex networks. *IEEE Control Systems Magazine*, 36(1):26–46, 2016.

[18] Elad Schneidman, Michael J Berry, Ronen Segev, and William Bialek. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440(7087):1007–1012, 2006.

[19] Shan Yu, Hongdian Yang, Hiroyuki Nakahara, Gustavo Santos, Danko Nikolić, and Dietmar Plenz. Higher-order interactions characterized in cortical activity. *J. Neurosci.*, 31:17514–26, 11 2011.

[20] Andrea Santoro, Matteo Neri, Simone Poetto, Davide Orsenigo, Matteo Diano, Marilyn Gatica, and GIovanni Petri. Beyond pairwise interactions: Charting higher-order models of brain function. *bioRxiv*, pages 2025–06, 2025.

[21] Staša Milojević. Principles of scientific research team formation and evolution. *Proceedings of the National Academy of Sciences*, 111(11):3984–3989, 2014.

[22] Damon Centola, Joshua Becker, Devon Brackbill, and Andrea Baronchelli. Experimental evidence for tipping points in social convention. *Science*, 360(6393):1116–1119, 2018.

[23] Jacopo Grilli, György Barabás, Matthew J Michalska-Smith, and Stefano Allesina. Higher-order interactions stabilize dynamics in competitive network models. *Nature*, 548(7666):210–213, 2017.

[24] Margaret Mayfield and Daniel Stouffer. Higher-order interactions capture unexplained complexity in diverse communities. *Nat. Ecol. Evol.*, 1:0062, 02 2017.

[25] Alicia Sanchez-Gorostiaga, Djordje Bajić, Melisa L Osborne, Juan F Poyatos, and Alvaro Sanchez. High-order interactions distort the functional landscape of microbial consortia. *PLoS Biol.*, 17(12):e3000550, 2019.

[26] Federico Battiston, Giulia Cencetti, Iacopo Iacopini, Vito Latora, Maxime Lucas, Alice Patania, Jean-Gabriel Young, and Giovanni Petri. Networks beyond pairwise interactions: structure and dynamics. *Phys. Rep.*, 874:1–92, 2020.

[27] Christian Bick, Elizabeth Gross, Heather A Harrington, and Michael T Schaub. What are higher-order networks? *SIAM Review*, 65(3):686–731, 2023.

[28] Yuanzhao Zhang, Maxime Lucas, and Federico Battiston. Higher-order interactions shape collective dynamics differently in hypergraphs and simplicial complexes. *Nat. Commun.*, 14(1):1605, 2023.

[29] Michael T Schaub and Santiago Segarra. Flow smoothing and denoising: Graph signal processing in the edge-space. In *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 735–739. IEEE, 2018.

[30] Sergio Barbarossa and Stefania Sardellitti. Topological signal processing over simplicial complexes. *IEEE Transactions on Signal Processing*, 68:2992–3007, 2020.

[31] Guilherme Ferraz de Arruda, Giovanni Petri, Pablo Martín Rodriguez, and Yamir Moreno. Multistability, intermittency, and hybrid transitions in social contagion models on hypergraphs. *Nat. Commun.*, 14(1):1375, 2023.

[32] Christian Kuehn and Christian Bick. A universal route to explosive phenomena. *Sci. Adv.*, 7(16):eabe3824, 2021.

[33] Ana P Millán, Joaquín J Torres, and Ginestra Bianconi. Explosive higher-order kuramoto dynamics on simplicial complexes. *Phys. Rev. Lett.*, 124(21):218301, 2020.

[34] Federico Battiston, Enrico Amico, Alain Barrat, Ginestra Bianconi, Guilherme Ferraz de Arruda, Benedetta Franceschiello, Iacopo Iacopini, Sonia Kéfi, Vito Latora, Yamir Moreno, et al. The physics of higher-order interactions in complex systems. *Nature Physics*, 17(10):1093–1098, 2021.

[35] Per Sebastian Skardal and Alex Arenas. Abrupt desynchronization and extensive multistability in globally coupled oscillator simplexes. *Physical Review Letters*, 122(24):248301, 2019.

[36] Per Sebastian Skardal, Lluís Arola-Fernández, Dane Taylor, and Alex Arenas. Higher-order interactions can better optimize network synchronization. *Phys. Rev. Res.*, 3(4):043193, 2021.

[37] Maxime Lucas, Giulia Cencetti, and Federico Battiston. Multiorder laplacian for synchronization in higher-order networks. *Phys. Rev. Res.*, 2(3):033410, 2020.

[38] Erik Gengel, Erik Teichmann, Michael Rosenblum, and Arkady Pikovsky. High-order phase reduction for coupled oscillators. *J. phys. Complex*, 2(1):015005, 2020.

[39] Matthew H Matheny, Jeffrey Emenheiser, Warren Fon, Airlie Chapman, Anastasiya Salova, Martin Rohden, Jarvis Li, Mathias Hudoba de Badyn, Márton Pósfai, Leonardo Duenas-Osorio, et al. Exotic states in a simple network of nanoelectromechanical oscillators. *Science*, 363(6431):eaav7932, 2019.

[40] Lucia Valentina Gambuzza, Francesca Di Patti, Luca Gallo, Stefano Lepri, Miguel Romance, Regino Criado, Mattia Frasca, Vito Latora, and Stefano Boccaletti. Stability of synchronization in simplicial complexes. *Nat. Commun.*, 12(1):1255, 2021.

[41] Srilena Kundu and Dibakar Ghosh. Higher-order interactions promote chimera states. *Phys. Rev. E*, 105(4):L042202, 2022.

[42] Christian Bick, Peter Ashwin, and Ana Rodrigues. Chaos in generically coupled phase oscillator networks with nonpairwise interactions. *Chaos*, 26(9):094814, 2016.

[43] Christian Bick. Heteroclinic dynamics of localized frequency synchrony: Heteroclinic cycles for small populations. *J. Nonlin. Sci.*, 29:2571–2600, 2019.

[44] Christian Bick and Alexander Lohse. Heteroclinic dynamics of localized frequency synchrony: Stability of heteroclinic cycles and networks. *J. Nonlin. Sci.*, 29:2547–2570, 2019.

[45] Sabina Adhikari, Juan G Restrepo, and Per Sebastian Skardal. Synchronization of phase oscillators on complex hypergraphs. *Chaos*, 33(3), 2023.

[46] K Kovalenko, X Dai, K Alfaro-Bittner, AM Raigorodskii, M Perc, and S Boccaletti. Contrarians synchronize beyond the limit of pairwise interactions. *Phys. Rev. Lett.*, 127(25):258301, 2021.

[47] Iván León and Diego Pazó. Phase reduction beyond the first order: The case of the mean-field complex Ginzburg-Landau equation. *Phys. Rev. E*, 100(1):012211, 2019.

[48] Chen Chris Gong and Arkady Pikovsky. Low-dimensional dynamics for higher-order harmonic, globally coupled phase-oscillator ensembles. *Phys. Rev. E*, 100(6):062210, 2019.

[49] Yuanzhao Zhang, Per Sebastian Skardal, Federico Battiston, Giovanni Petri, and Maxime Lucas. Deeper but smaller: Higher-order interactions increase linear stability but shrink basins. *Science Advances*, 10(40):eado8049, 2024.

[50] Alexis Arnaudon, Robert L Peach, Giovanni Petri, and Paul Expert. Connecting hodge and sakaguchi-kuramoto through a mathematical framework for coupled oscillators on simplicial complexes. *Commun. Phys*, 5(1):1–12, 2022.

[51] Lucille Calmon, Michael T Schaub, and Ginestra Bianconi. Dirac signal processing of higher-order topological signals. *New Journal of Physics*, 25(9):093013, 2023.

[52] Abubakr Muhammad and Magnus Egerstedt. Control using higher order Laplacians in network topologies. In *Proc. of 17th International Symposium on Mathematical Theory of Networks and Systems*, pages 1024–1038. Citeseer, 2006.

[53] Ana P Millán, Hanlin Sun, Lorenzo Giambagli, Riccardo Muolo, Timoteo Carletti, Joaquín J Torres, Filippo Radicchi, Jürgen Kurths, and Ginestra Bianconi. Topology shapes dynamics of higher-order networks. *Nature Physics*, pages 1–9, 2025.

[54] Marco Nurisso, Alexis Arnaudon, Maxime Lucas, Robert L Peach, Paul Expert, Francesco Vaccarino, and Giovanni Petri. A unified framework for simplicial kuramoto models. *Chaos*, 34(5), 2024.

[55] Michael E Fisher. The renormalization group in the theory of critical behavior. *Reviews of Modern Physics*, 46(4):597, 1974.

[56] Chaoming Song, Shlomo Havlin, and Hernan A Makse. Self-similarity of complex networks. *Nature*, 433(7024):392–395, 2005.

[57] K-I Goh, Giovanni Salvi, Byungnam Kahng, and Doochul Kim. Skeleton and fractal scaling in complex networks. *Physical Review Letters*, 96(1):018701, 2006.

[58] Hernán D Rozenfeld, Chaoming Song, and Hernán A Makse. Small-world to fractal transition in complex networks: A renormalization group approach. *Physical Review Letters*, 104(2):025701, 2010.

[59] Guillermo García-Pérez, Marián Boguñá, and M Ángeles Serrano. Multiscale unfolding of real networks by geometric renormalization. *Nature Physics*, 14(6):583–589, 2018.

[60] Elena Garuccio, Margherita Lalli, and Diego Garlaschelli. Multiscale network renormalization: Scale-invariance without geometry. *Physical Review Research*, 5(4):043101, 2023.

[61] Muhua Zheng, Guillermo García-Pérez, Marián Boguñá, and M Ángeles Serrano. Geometric renormalization of weighted networks. *Communications Physics*, 7(1):97, 2024.

[62] Matheus de C. Loures, Alan Albert Piovesana, and Joséé Antôônio Brum. Laplacian coarse graining in complex networks. *arXiv*, February 2023.

[63] Pablo Villegas, Tommaso Gili, Guido Caldarelli, and Andrea Gabrielli. Laplacian renormalization group for heterogeneous networks. *Nature Physics*, 19(3):445–450, 2023.

[64] Yuhai Tu. The renormalization group for non-equilibrium systems. *Nature Physics*, 19(11):1536–1538, 2023.

[65] M Ángeles Serrano, Dmitri Krioukov, and Marián Boguná. Self-similarity of complex networks and hidden metric spaces. *Physical Review Letters*, 100(7):078701, 2008.

[66] Pablo Villegas, Andrea Gabrielli, Francesca Santucci, Guido Caldarelli, and Tommaso Gili. Laplacian paths in complex networks: Information core emerges from entropic transitions. *Physical Review Research*, 4(3):033196, 2022.

[67] Marco Nurisso, Marta Morandini, Maxime Lucas, Francesco Vaccarino, Tommaso Gili, and Giovanni Petri. Higher-order laplacian renormalization. *Nature Physics*, pages 1–8, 2025.

[68] Guido Caldarelli, Andrea Gabrielli, Tommaso Gili, and Pablo Villegas. Laplacian renormalization group: an introduction to heterogeneous coarse-graining. *Journal of Statistical Mechanics: Theory and Experiment*, 2024(8):084002, 2024.

[69] Michael Breakspear, Stewart Heitmann, and Andreas Daffertshofer. Generative models of cortical oscillations: neurobiological implications of the kuramoto model. *Front. Hum. Neurosci.*, 4:190, 2010.

[70] Steven H Strogatz. Spontaneous synchronization in nature. In *Proceedings of international frequency control symposium*, pages 2–4. IEEE, 1997.

[71] Martin Rohden, Andreas Sorge, Marc Timme, and Dirk Witthaut. Self-organized synchronization in decentralized power grids. *Phys. Rev. Lett.*, 109(6):064101, 2012.

[72] Takashi Nishikawa and Adilson E Motter. Comparative analysis of existing models for power-grid synchronization. *New J. Phys.*, 17(1):015012, 2015.

[73] Zoltán Néda, Erzsébet Ravasz, Tamás Vicsek, Yves Brechet, and Albert-Lázló Barabási. Physics of the rhythmic applause. *Phys. Rev. E*, 61(6):6987, 2000.

[74] Yoshiki Kuramoto. Self-entrainment of a population of coupled non-linear oscillators. In Huzihiro Araki, editor, *International Symposium on Mathematical Problems in Theoretical Physics*, pages 420–422, Berlin, Heidelberg, 1975. Springer Berlin Heidelberg.

[75] A. Arenas, A. Díaz-Guilera, J. Kurths, Y. Moreno, and C. Zhou. Synchronization in complex networks. *Phys. Rep.*, 469(3):93–153, 2008.

[76] A. Jadbabaie, N. Motee, and M. Barahona. On the stability of the kuramoto model of coupled nonlinear oscillators. In *Proceedings of the 2004 American Control Conference*, volume 5, pages 4296–4301 vol.5, 2004.

[77] Federico Battiston, Enrico Amico, Alain Barrat, Ginestra Bianconi, Guilherme Ferraz de Arruda, Benedetta Franceschiello, Iacopo Iacopini, Sonia Kéfi, Vito Latora, Yamir Moreno, Micah M. Murray, Tiago P. Peixoto, Francesco Vaccarino, and Giovanni Petri. The physics of higher-order interactions in complex systems. *Nat. Phys.*, 17(10):1093–1098, oct 2021.

[78] Christian Bick, Elizabeth Gross, Heather A Harrington, and Michael T Schaub. What are higher-order networks? *SIAM Review*, 65(3):686–731, 2023.

[79] Michael T. Schaub, Austin R. Benson, Paul Horn, Gabor Lippner, and Ali Jadbabaie. Random walks on simplicial complexes and the normalized hodge 1-laplacian. *SIAM Review*, 62(2):353–391, 2020.

[80] Timoteo Carletti, Federico Battiston, Giulia Cencetti, and Duccio Fanelli. Random walks on hypergraphs. *Phys. Rev. E*, 101(2):022308, 2020.

[81] Ana P Millán, Reza Ghorbanchian, Nicolò Defenu, Federico Battiston, and Ginestra Bianconi. Local topological moves determine global diffusion properties of hyperbolic higher-order networks. *Phys. Rev. E*, 104(5):054302, 2021.

[82] Leonie Neuhäuser, Andrew Mellor, and Renaud Lambiotte. Multibody interactions and nonlinear consensus dynamics on networked systems. *Phys. Rev. E*, 101(3):032310, March 2020.

[83] Lee DeVille. Consensus on simplicial complexes: Results on stability and synchronization. *Chaos*, 31(2):023137, 2021.

[84] Iacopo Iacopini, Giovanni Petri, Andrea Baronchelli, and Alain Barrat. Group interactions modulate critical mass dynamics in social convention. *Commun. Phys*, 5(1):64, 2022.

[85] Maxime Lucas, Iacopo Iacopini, Thomas Robiglio, Alain Barrat, and Giovanni Petri. Simplicially driven simple contagion. *Phys. Rev. Res.*, 5(1):013201, 2023.

[86] Iacopo Iacopini, Giovanni Petri, Alain Barrat, and Vito Latora. Simplicial models of social contagion. *Nat. Commun.*, 10(1):2485, 2019.

[87] Guilherme Ferraz de Arruda, Michele Tizzani, and Yamir Moreno. Phase transitions and stability of dynamical processes on hypergraphs. *Commun. Phys.*, 4:24, 2021.

[88] Sandeep Chowdhary, Aanjaneya Kumar, Giulia Cencetti, Iacopo Iacopini, and Federico Battiston. Simplicial contagion in temporal higher-order networks. *J. Phys. Complex*, 2(3):035019, September 2021.

[89] Guillaume St-Onge, Hanlin Sun, Antoine Allard, Laurent Hébert-Dufresne, and Ginestra Bianconi. Universal nonlinear infection kernel from heterogeneous exposure on higher-order networks. *Phys. Rev. Lett.*, 127(15):158301, 2021.

[90] Guillaume St-Onge, Vincent Thibeault, Antoine Allard, Louis J Dubé, and Laurent Hébert-Dufresne. Master equation analysis of mesoscopic localization in contagion dynamics on higher-order networks. *Phys. Rev. E*, 103(3):032301, 2021.

[91] Guillaume St-Onge, Iacopo Iacopini, Vito Latora, Alain Barrat, Giovanni Petri, Antoine Allard, and Laurent Hébert-Dufresne. Influential groups for seeding and sustaining nonlinear contagion in heterogeneous hypergraphs. *Commun. Phys*, 5(1):25, 2022.

[92] Hanlin Sun, Filippo Radicchi, Jürgen Kurths, and Ginestra Bianconi. The dynamic nature of percolation on networks with triadic interactions. *Nat. Commun.*, 14(1):1308, 2023.

[93] Ginestra Bianconi, Ivan Kryven, and Robert M Ziff. Percolation on branching simplicial and cell complexes and its relation to interdependent percolation. *Phys. Rev. E*, 100(6):062311, 2019.

[94] Hanlin Sun, Robert M Ziff, and Ginestra Bianconi. Renormalization group theory of percolation on pseudofractal simplicial and cell complexes. *Phys. Rev. E*, 102(1):012308, 2020.

[95] Unai Alvarez-Rodriguez, Federico Battiston, Guilherme Ferraz de Arruda, Yamir Moreno, Matjaž Perc, and Vito Latora. Evolutionary dynamics of higher-order interactions in social networks. *Nat. Hum. Behav*, 5(5):586–595, 2021.

[96] Lucille Calmon, Juan G Restrepo, Joaquín J Torres, and Ginestra Bianconi. Dirac synchronization is rhythmic and explosive. *Commun. Phys*, 5(1):1–17, 2022.

[97] Lucille Calmon, Sanjukta Krishnagopal, and Ginestra Bianconi. Local dirac synchronization on networks. *Chaos*, 33(3):033117, 2023.

[98] Ginestra Bianconi. The topological dirac equation of networks and simplicial complexes. *J. phys. Complex*, 2(3):035022, sep 2021.

[99] Juan A. Acebrón, L. L. Bonilla, Conrad J. Pérez Vicente, Félix Ritort, and Renato Spigler. The kuramoto model: A simple paradigm for synchronization phenomena. *Rev. Mod. Phys.*, 77:137–185, Apr 2005.

[100] Arkady Pikovsky and Michael Rosenblum. Dynamics of globally coupled oscillators: Progress and perspectives. *Chaos*, 25(9), 2015.

[101] Francisco A Rodrigues, Thomas K DM Peron, Peng Ji, and Jürgen Kurths. The kuramoto model in complex networks. *Physics Reports*, 610:1–98, 2016.

[102] Eddie Nijholt and Lee DeVille. Dynamical systems defined on simplicial complexes: Symmetries, conjugacies, and invariant subspaces. *Chaos*, 32(9):093131, sep 2022.

[103] Timoteo Carletti, Lorenzo Giambagli, and Ginestra Bianconi. Global topological synchronization on simplicial and cell complexes. *Phys. Rev. Lett.*, 130(18):187401, 2023.

[104] Hiroaki Daido. Generic scaling at the onset of macroscopic mutual entrainment in limit-cycle oscillators with uniform all-to-all coupling. *Phys. Rev. Lett.*, 73(5):760, 1994.

[105] David Hansel, German Mato, and Claude Meunier. Clustering and slow switching in globally coupled phase oscillators. *Phys. Rev. E*, 48(5):3470, 1993.

[106] Hiroaki Daido. Order function and macroscopic mutual entrainment in uniformly coupled limit-cycle oscillators. *Progress of theoretical physics*, 88(6):1213–1218, 1992.

[107] Hiroaki Daido. Critical conditions of macroscopic mutual entrainment in uniformly coupled limit-cycle oscillators. *Progress of theoretical physics*, 89(4):929–934, 1993.

[108] Hiroaki Daido. A solvable model of coupled limit-cycle oscillators exhibiting partial perfect synchrony and novel frequency spectra. *Physica D: Nonlinear Phenomena*, 69(3-4):394–403, 1993.

[109] H Daido. Multi-branch entrainment and multi-peaked order-functions in a phase model of limit-cycle oscillators with uniform all-to-all coupling. *Journal of Physics A: Mathematical and General*, 28(5):L151, 1995.

[110] Hiroaki Daido. Multibranch entrainment and scaling in large populations of coupled oscillators. *Phys. Rev. Lett.*, 77(7):1406, 1996.

[111] Hiroaki Daido. Onset of cooperative entrainment in limit-cycle oscillators with uniform all-to-all interactions: bifurcation of the order function. *Physica D: Nonlinear Phenomena*, 91(1-2):24–66, 1996.

[112] Peter Ashwin, Gábor Orosz, John Wordsworth, and Stuart Townley. Dynamics on networks of cluster states for globally coupled phase oscillators. *SIAM Journal on Applied Dynamical Systems*, 6(4):728–758, 2007.

[113] M Komarov and A Pikovsky. The kuramoto model of coupled oscillators with a bi-harmonic coupling function. *Physica D: Nonlinear Phenomena*, 289:18–31, 2014.

[114] Maxim Komarov and Arkady Pikovsky. Dynamics of multifrequency oscillator communities. *Phys. Rev. Lett.*, 110(13):134101, 2013.

[115] VK Chandrasekar, M Manoranjani, and Shamik Gupta. Kuramoto model in the presence of additional interactions that break rotational symmetry. *Phys. Rev. E*, 102(1):012206, 2020.

[116] L. Eldén. A weighted pseudoinverse, generalized singular values, and constrained least squares problems. *BIT*, 22:487–502, 01 1982.

[117] Debsankha Manik, Marc Timme, and Dirk Witthaut. Cycle flows and multistability in oscillatory networks. *Chaos*, 27(8):083123, 08 2017.

[118] Florian Dörfler and Francesco Bullo. On the critical coupling for kuramoto oscillators. *SIAM J. Appl. Dyn. Syst.*, 10(3):1070–1099, 2011.

[119] Florian Dörfler and Francesco Bullo. Synchronization in complex networks of phase oscillators: A survey. *Automatica*, 50(6):1539–1564, 2014.

[120] Florian Dörfler and Francesco Bullo. Exploring synchronization in complex oscillator networks. In *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, pages 7157–7170. IEEE, 2012.

[121] Mitchell Black and William Maxwell. Effective Resistance and Capacitance in Simplicial Complexes and a Quantum Algorithm. In Hee-Kap Ahn and Kunihiko Sadakane, editors, *32nd International Symposium on Algorithms and Computation (ISAAC 2021)*, volume 212 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 31:1–31:27, Dagstuhl, Germany, 2021. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

[122] Morten L Kringelbach and Gustavo Deco. Brain states and transitions: insights from computational neuroscience. *Cell Reports*, 32(10), 2020.

[123] Gyorgy Buzsaki. *Rhythms of the Brain*. Oxford university press, 2006.

[124] Louis-David Lord, Timoteo Carletti, Henrique Fernandes, Federico E. Turkheimer, and Paul Expert. Altered dynamical integration/segregation balance during anesthesia-induced loss of consciousness. *Frontiers in Network Physiology*, 3:1279646, 2023.

[125] Joana Cabral, Etienne Hugues, Olaf Sporns, and Gustavo Deco. Role of local network oscillations in resting-state functional connectivity. *NeuroImage*, 57(1):130–139, 2011.

[126] Louis-David Lord, Paul Expert, Jeremy F Huckins, and Federico E Turkheimer. Cerebral energy metabolism and the brain's functional network architecture: an integrative review. *Journal of Cerebral Blood Flow & Metabolism*, 33(9):1347–1354, 2013.

[127] David Cumin and CP2296240 Unsworth. Generalising the kuramoto model for the study of neuronal synchronisation in the brain. *Phys. D*, 226(2):181–196, 2007.

[128] Ruben Schmidt, Karl JR LaFleur, Marcel A de Reus, Leonard H van den Berg, and Martijn P van den Heuvel. Kuramoto model simulation of neural hubs and dynamic synchrony in the human cerebral connectome. *BMC Neurosci.*, 16(1):1–13, 2015.

[129] Maria Pope, Makoto Fukushima, Richard F Betzel, and Olaf Sporns. Modular origins of high-amplitude cofluctuations in fine-scale functional connectivity dynamics. *Proc. Natl. Acad. Sci. U.S.A.*, 118(46):e2109380118, 2021.

[130] Maria Pope, Caio Seguin, Thomas F Varley, Joshua Faskowitz, and Olaf Sporns. Co-evolving dynamics and topology in a coupled oscillator model of resting brain function. *bioRxiv*, pages 2023–01, 2023.

[131] František Váša, Murray Shanahan, Peter J Hellyer, Gregory Scott, Joana Cabral, and Robert Leech. Effects of lesions on synchrony and metastability in cortical networks. *Neuroimage*, 118:456–467, 2015.

[132] Shan Yu, Hongdian Yang, Hiroyuki Nakahara, Gustavo S Santos, Danko Nikolić, and Dietmar Plenz. Higher-order interactions characterized in cortical activity. *J. Neurosci.*, 31(48):17514–17526, 2011.

[133] Reza Ghorbanchian, Juan G Restrepo, Joaquín J Torres, and Ginestra Bianconi. Higher-order simplicial synchronization of coupled topological signals. *Commun. Phys*, 4(1):120, 2021.

[134] Joshua Faskowitz, Richard F Betzel, and Olaf Sporns. Edges in brain networks: Contributions to models of structure and function. *Netw. Neurosci.*, 6(1):1–28, 2022.

[135] Albert Gidon, Timothy Adam Zolnik, Pawel Fidzinski, Felix Bolduan, Athanasia Papoutsi, Panayiota Poirazi, Martin Holtkamp, Imre Vida, and Matthew Evan Larkum. Dendritic action potentials and computation in human layer 2/3 cortical neurons. *Science*, 367(6473):83–87, 2020.

[136] Andrea Santoro, Marco Nurisso, and Giovanni Petri. From nodes to edges: Edge-based laplacians for brain signal processing. *arXiv preprint arXiv:2512.13420*, 2025.

[137] Darrell Haufler and Denis Paré. Detection of multiway gamma coordination reveals how frequency mixing shapes neural dynamics. *Neuron*, 101(4):603–614, 2019.

[138] Charlotte Luff, Robert L Peach, Emma-Jane Mallas, Edward Rhodes, Felix Laumann, Edward S Boyden, David J Sharp, Mauricio Barahona, and Nir Grossman. The neuron mixer and its impact on human brain dynamics. *bioRxiv*, pages 2023–01, 2023.

[139] Iva Bačić, Michael T Schaub, Jürgen Kurths, and Dirk Witthaut. Phase locking and multistability in the topological kuramoto model on cell complexes. *arXiv preprint arXiv:2510.05831*, 2025.

[140] Riccardo Muolo, Timoteo Carletti, and Ginestra Bianconi. The three way dirac operator and dynamical turing and dirac induced patterns on nodes and links. *Chaos, Solitons & Fractals*, 178:114312, 2024.

[141] Ahmed AA Zaid and Ginestra Bianconi. Designing topological cluster synchronization patterns with the dirac operator. *arXiv preprint arXiv:2507.20837*, 2025.

[142] Leo P Kadanoff. Notes on migdal's recursion formulas. *Annals of Physics*, 100(1-2):359–394, 1976.

[143] Andrea Gabrielli, Diego Garlaschelli, Subodh P Patil, and M Ángeles Serrano. Network renormalization. *Nature Reviews Physics*, pages 1–17, 2025.

[144] Jin Seop Kim, Kwang-Il Goh, Byungnam Kahng, and Doochul Kim. Fractality and self-similarity in scale-free networks. *New Journal of Physics*, 9(6):177, 2007.

[145] Agata Fronczak, Piotr Fronczak, Mateusz J Samsel, Kordian Makulski, Michał Łepek, and Maciej J Mrowinski. Scaling theory of fractal complex networks. *Scientific Reports*, 14(1):9079, 2024.

[146] Jasper Van Der Kolk, Marián Boguñá, and M Ángeles Serrano. Renormalization of networks with weak geometric coupling. *Physical Review E*, 110(3):L032302, 2024.

[147] Leo P Kadanoff. Scaling laws for ising models near T c. *Physics Physique Fizika*, 2(6):263, 1966.

[148] Marian Boguna, Ivan Bonamassa, Manlio De Domenico, Shlomo Havlin, Dmitri Krioukov, and M Ángeles Serrano. Network geometry. *Nature Reviews Physics*, 3(2):114–135, 2021.

[149] Naoki Masuda, Mason A Porter, and Renaud Lambiotte. Random walks and diffusion on networks. *Physics reports*, 716:1–58, 2017.

[150] Ginestra Bianconi and Sergey N Dorogovstev. The spectral dimension of simplicial complexes: A renormalization group theory. *Journal of Statistical Mechanics: Theory and Experiment*, 2020(1):014005, 2020.

[151] Marcus Reitz and Ginestra Bianconi. The higher-order spectrum of simplicial complexes: A renormalization group approach. *J. Phys. A Math. Theor.*, 53(29):295001, 2020.

[152] Manlio De Domenico and Jacob Biamonte. Spectral entropies as information-theoretic tools for complex network comparison. *Physical Review X*, 6(4):041062, 2016.

[153] Arsham Ghavasieh, Carlo Nicolini, and Manlio De Domenico. Statistical physics of complex information dynamics. *Phys. Rev. E*, 102:052304, Nov 2020.

[154] Aohua Cheng, Yunhui Xu, Pei Sun, and Yang Tian. A simplex path integral and a simplex renormalization group for high-order interactions. *Reports on Progress in Physics*, 87(8):087601, 2024.

[155] Lek-Heng Lim. Hodge laplacians on graphs. *Siam Review*, 62(3):685–715, 2020.

[156] Sanjukta Krishnagopal and Ginestra Bianconi. Spectral detection of simplicial communities via hodge laplacians. *Physical Review E*, 104(6):064303, 2021.

[157] Forman. Bochner's method for cell complexes and combinatorial Ricci curvature. *Discrete & Computational Geometry*, 29:323–374, 2003.

[158] Mustafa Hajij, Ghada Zamzmi, Theodore Papamarkou, Nina Miolane, Aldo Guzmán-Sáenz, Karthikeyan Natesan Ramamurthy, Tolga Birdal, Tamal K Dey, Soham Mukherjee, Shreyas N Samaga, et al. Topological deep learning: Going beyond graph data. *arXiv preprint*, 2023.

[159] Mehmet Emin Aktas and Esra Akbas. Hypergraph laplacians in diffusion framework. In *Complex Networks & Their Applications X: Volume 2, Proceedings of the Tenth International Conference on Complex Networks and Their Applications COMPLEX NETWORKS 2021 10*, pages 277–288. Springer, 2022.

[160] Ernesto Estrada and Grant J Ross. Centralities in simplicial complexes. Applications to protein interaction networks. *Journal of theoretical biology*, 438:46–60, 2018.

[161] Ronald H Atkin. From cohomology in physics to q-connectivity in social science. *International journal of man-machine studies*, 4(2):139–167, 1972.

[162] RH Atkin. An algebra for patterns on a complex, ii. *International Journal of Man-Machine Studies*, 8(5):483–498, 1976.

[163] Ronald Harry Atkin. *Mathematical structure in human affairs*. Heinemann Educational London, 1974.

[164] Hassler Whitney. Congruent graphs and the connectivity of graphs. *Hassler Whitney Collected Papers*, pages 61–79, 1992.

[165] Federica Baccini, Filippo Geraci, and Ginestra Bianconi. Weighted simplicial complexes and their representation power of higher-order network data and topology. *Physical Review E*, 106(3):034319, 2022.

[166] A Ghavasieh and Manlio De Domenico. Statistical physics of network structure and information dynamics. *Journal of Physics: Complexity*, 3(1):011001, 2022.

[167] Michael A Nielsen and Isaac L Chuang. *Quantum computation and quantum information*. Cambridge university press, 2010.

[168] Anna Poggialini, Pablo Villegas, Miguel A Muñoz, and Andrea Gabrielli. Networks with many structural scales: A renormalization group perspective. *Physical Review Letters*, 134(5):057401, 2025.

[169] Arsham Ghavasieh and Manlio De Domenico. Generalized network density matrices for analysis of multiscale functional diversity. *Phys. Rev. E*, 107:044304, Apr 2023.

[170] Arsham Ghavasieh and Manlio De Domenico. Diversity of information pathways drives sparsity in real-world networks. *Nature Physics*, pages 1–8, 2024.

[171] Rammal Rammal and Gérard Toulouse. Random walks on fractal structures and percolation clusters. *Journal de Physique Lettres*, 44(1):13–22, 1983.

[172] Jan Ambjørn, Jerzy Jurkiewicz, and Renate Loll. The spectral dimension of the universe is scale dependent. *Physical Review Letters*, 95(17):171301, 2005.

[173] Gianluca Calcagni, Daniele Oriti, and Johannes Thürigen. Spectral dimension of quantum geometries. *Classical and Quantum Gravity*, 31(13):135014, 2014.

[174] Ana P Millán, Joaquín J Torres, and Ginestra Bianconi. Synchronization in network geometries with finite spectral dimension. *Physical Review E*, 99(2):022307, 2019.

[175] Vincent Peter Grande and Michael T Schaub. Topological point cloud clustering. In *International Conference on Machine Learning*, pages 11683–11697. PMLR, 2023.

[176] James R Munkres. *Elements of algebraic topology*. CRC press, 2018.

[177] Sergey N Dorogovtsev, Alexander V Goltsev, and José Ferreira F Mendes. Pseudofractal scale-free web. *Physical Review E*, 65(6):066122, 2002.

[178] Ginestra Bianconi and Christoph Rahmede. Network geometry with flavor: From complexity to quantum geometry. *Physical Review E*, 93(3):032315, 2016.

[179] Ginestra Bianconi and Christoph Rahmede. Emergent hyperbolic network geometry. *Scientific reports*, 7(1):41974, 2017.

[180] Rossana Mastrandrea, Julie Fournet, and Alain Barrat. Contact Patterns in a High School: A Comparison between Data Collected Using Wearable Sensors, Contact Diaries and Friendship Surveys. *PLoS One*, 10(9):e0136497, September 2015.

[181] Kwang-Il Goh, Michael E Cusick, David Valle, Barton Childs, Marc Vidal, and Albert-László Barabási. The human disease network. *Proceedings of the National Academy of Sciences of the United States of America*, 104(21):8685–8690, 2007.

[182] Austin R. Benson, Rediet Abebe, Michael T. Schaub, Ali Jadbabaie, and Jon Kleinberg. Simplicial closure and higher-order link prediction. *Proc Natl Acad Sci U.S.A*, 115(48):E11221–E11230, November 2018.

[183] Enron email dataset, July 2024.

[184] Lovro Šubelj and Marko Bajec. Robust network community detection using balanced propagation. *The European Physical Journal B*, 81:353–362, 2011.

[185] Mark EJ Newman. Finding community structure in networks using the eigenvectors of matrices. *Physical Review E*, 74(3):036104, 2006.

[186] Roger Guimera, Leon Danon, Albert Diaz-Guilera, Francesc Giralt, and Alex Arenas. Self-similar community structure in a network of human interactions. *Physical Review E*, 68(6):065103, 2003.

[187] Jordi Duch and Alex Arenas. Community detection in complex networks using extremal optimization. *Physical Review E*, 72(2):027104, 2005.

[188] Hawoong Jeong, Bálint Tombor, Réka Albert, Zoltan N Oltvai, and A-L Barabási. The large-scale organization of metabolic networks. *Nature*, 407(6804):651–654, 2000.

[189] Ross Overbeek, Niels Larsen, Gordon D Pusch, Mark D'Souza, Evgeni Selkov Jr, Nikos Kyrpides, Michael Fonstein, Natalia Maltsev, and Evgeni Selkov. WIT: Integrated system for high-throughput genome sequence analysis and metabolic reconstruction. *Nucleic acids research*, 28(1):123–125, 2000.

[190] Andrew Beveridge and Michael Hunger. Asoiaf, 2013.

[191] Giovanni Petri, Martina Scolamiero, Irene Donato, and Francesco Vaccarino. Topological strata of weighted complex networks. *PloS one*, 8(6):e66506, 2013.

[192] Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, et al. *An introduction to statistical learning*, volume 112. Springer, 2013.

[193] Vincent Thibeault, Antoine Allard, and Patrick Desrosiers. The low-rank hypothesis of complex systems. *Nature Physics*, pages 1–9, 2024.

[194] Hanbaek Lyu, Yacoub H Kureh, Joshua Vendrow, and Mason A Porter. Learning low-rank latent mesoscale structures in networks. *Nature Communications*, 15(1):224, 2024.

[195] Guilherme Ferraz de Arruda, Giovanni Petri, Pablo Martin Rodriguez, and Yamir Moreno. Multistability, intermittency, and hybrid transitions in social contagion models on hypergraphs. *Nature Communications*, 14(1):1375, 2023.

[196] Iacopo Iacopini, Giovanni Petri, Alain Barrat, and Vito Latora. Simplicial models of social contagion. *Nature Communications*, 10:2485, June 2019.

[197] Maxime Lucas, Luca Gallo, Arsham Ghavasieh, Federico Battiston, and Manlio De Domenico. Reducibility of higher-order networks via information flow. *arXiv preprint arXiv:2404.08547*, 2024.

[198] Charles Murphy, Vincent Thibeault, Antoine Allard, and Patrick Desrosiers. Duality between predictability and reconstructability in complex systems. *Nature Communications*, 15(1):4478, 2024.

[199] Juliette Stehlé, Nicolas Voirin, Alain Barrat, Ciro Cattuto, Lorenzo Isella, Jean-François Pinton, Marco Quaggiotto, Wouter Van den Broeck, Corinne Régis, Bruno Lina, and Philippe Vanhems. High-resolution measurements of face-to-face contact patterns in a primary school. *PLoS One*, 6(8):e23176, August 2011.

[200] Valerio Gemmetto, Alain Barrat, and Ciro Cattuto. Mitigation of infectious disease at school: targeted class closure vs school closure. *BMC Infect Dis*, 14(1):1–10, December 2014.

[201] The sociopatterns collaboration.

[202] Hao Yin, Austin R. Benson, Jure Leskovec, and David F. Gleich. Local higher-order graph clustering. In *ACM Conferences*, pages 555–564. Association for Computing Machinery, New York, NY, USA, August 2017.

[203] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM Trans Knowl Discov Data*, 1(1):2–es, March 2007.

[204] Philippe Vanhems, Alain Barrat, Ciro Cattuto, Jean-François Pinton, Nagham Khanafer, Corinne Régis, Byeul-a. Kim, Brigitte Comte, and Nicolas Voirin. Estimating potential infection transmission routes in hospital wards using wearable proximity sensors. *PLoS One*, 8(9):e73970, September 2013.

[205] Lorenzo Isella, Juliette Stehlé, Alain Barrat, Ciro Cattuto, Jean-François Pinton, and Wouter Van den Broeck. What's in a crowd? Analysis of face-to-face behavioral networks. *J Theor Biol*, 271(1):166–180, February 2011.

[206] Mathieu Génois, Christian L. Vestergaard, Julie Fournet, André Panisson, Isabelle Bonmarin, and Alain Barrat. Data on face-to-face contacts in an office building suggest a low-cost vaccination strategy based on community linkers. *Network Sci*, 3(3):326–347, September 2015.

[207] Mathieu Génois and Alain Barrat. Can co-location be used as a proxy for face-to-face contacts? *EPJ Data Sci*, 7(1):1–18, December 2018.

[208] Laura Ozella, Daniela Paolotti, Guilherme Lichand, Jorge P. Rodríguez, Simon Haenni, John Phuka, Onicio B. Leal-Neto, and Ciro Cattuto. Using wearable proximity sensors to characterize social contact patterns in a village of rural Malawi. *EPJ Data Sci*, 10(1):1–17, December 2021.

[209] Ciro Cattuto, Wouter Van den Broeck, Alain Barrat, Vittoria Colizza, Jean-François Pinton, and Alessandro Vespignani. Dynamics of Person-to-Person Interactions from Distributed RFID Sensor Networks. *PLoS One*, 5(7):e11596, July 2010.

[210] James H. Fowler. Connecting the Congress: A Study of Cosponsorship Networks. *Political Analysis*, 14(4):456–487, October 2006.

[211] James H. Fowler. Legislative cosponsorship networks in the US House and Senate. *Social Networks*, 28(4):454–465, October 2006.

[212] Janet Piñero, Juan Manuel Ramírez-Anguita, Josep Saüch-Pitarch, Francesco Ronzano, Emilio Centeno, Ferran Sanz, and Laura I. Furlong. The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res*, 48(D1):D845–D855, January 2020.

[213] Jérôme Kunegis. KONECT – The Koblenz Network Collection. In *Proc. Int. Conf. on World Wide Web Companion*, pages 1343–1350, 2013.

[214] Aaron Clauset, Ellen Tucker, and Matthias Sainz. The colorado index of complex networks. 2016.

[215] Ryan A. Rossi and Nesreen K. Ahmed. The network data repository with interactive graph analytics and visualization. In *AAAI*, 2015.

[216] Jure Leskovec and Julian Mcauley. Learning to discover social circles in ego networks. *Advances in neural information processing systems*, 25, 2012.

[217] Jean-François Rual, Kavitha Venkatesan, Tong Hao, Tomoko Hirozane-Kishikawa, Amélie Dricot, Ning Li, Gabriel F Berriz, Francis D Gibbons, Matija Dreze, Nono Ayivi-Guedehoussou, et al. Towards a proteome-scale map of the human protein–protein interaction network. *Nature*, 437(7062):1173–1178, 2005.

[218] Hawoong Jeong, Sean P Mason, A-L Barabási, and Zoltan N Oltvai. Lethality and centrality in protein networks. *Nature*, 411(6833):41–42, 2001.

[219] Stéphane Coulomb, Michel Bauer, Denis Bernard, and Marie-Claude Marsolier-Kergoat. Gene essentiality and the topology of protein interaction networks. *Proc. R. Soc. B: Biol. Sci.*, 272(1573):1721–1725, 2005.

[220] Jing-Dong J Han, Denis Dupuy, Nicolas Bertin, Michael E Cusick, and Marc Vidal. Effect of sampling on topology predictions of protein-protein interaction networks. *Nature Biotechnology*, 23(7):839–844, 2005.

[221] Michael PH Stumpf, Carsten Wiuf, and Robert M May. Subnets of scale-free networks are not scale-free: Sampling properties of networks. *Proceedings of the National Academy of Sciences of the United States of America*, 102(12):4221–4224, 2005.

[222] Marcus Kaiser and Claus C Hilgetag. Spatial growth of real-world networks. *Physical Review E*, 69(3):036103, 2004.

[223] Marcel A de Reus and Martijn P van den Heuvel. Rich club organization and intermodule communication in the cat connectome. *Journal of Neuroscience*, 33(32):12929–12939, 2013.

[224] Nikola T Markov, Julien Vezoli, Pascal Chameau, Arnaud Falchier, René Quilodran, Cyril Huissoud, Camille Lamy, Pierre Misery, Pascale Giroud, Shimon Ullman, et al. Anatomy of hierarchy: Feedforward and feedback pathways in macaque visual cortex. *Journal of Comparative Neurology*, 522(1):225–259, 2014.

[225] Davi D Bock, Wei-Chung Allen Lee, Aaron M Kerlin, Mark L Andermann, Greg Hood, Arthur W Wetzel, Sergey Yurgenson, Edward R Soucy, Hyon Suk Kim, and R Clay Reid. Network anatomy and in vivo physiology of visual cortical neurons. *Nature*, 471(7337):177–182, 2011.

[226] Travis A Jarrell, Yi Wang, Adam E Bloniarz, Christopher A Brittin, Meng Xu, J Nichol Thomson, Donna G Albertson, David H Hall, and Scott W Emmons. The connectome of a decision-making neural network. *Science (New York, N.Y.)*, 337(6093):437–444, 2012.

[227] N Blagus and M Bajec. The network of collaboration: Informatica and uporabna informatika. *Uporabna Informatika*, 23(1):22–31, 2015.

[228] Andrew Kachites McCallum, Kamal Nigam, Jason Rennie, and Kristie Seymore. Automating the construction of internet portals with machine learning. *Information Retrieval*, 3:127–163, 2000.

[229] C Dieter Zander, Neri Josten, Kim C Detloff, Robert Poulin, John P McLaughlin, and David W Thieltges. Food web including metazoan parasites for a brackish shallow water ecosystem in Germany and Denmark: Ecological Archives E092-174. *Ecology*, 92(10):2007–2007, 2011.

[230] E. C. Sancho. The hessen asymmetric network, 2016.

[231] D. Gleich. Minnesota road network, 2010.

[232] Dongbo Bu, Yi Zhao, Lun Cai, Hong Xue, Xiaopeng Zhu, Hongchao Lu, Jingfen Zhang, Shiwei Sun, Lunjiang Ling, Nan Zhang, et al. Topological structure analysis of the protein–protein interaction network in budding yeast. *Nucleic acids research*, 31(9):2443–2450, 2003.

[233] B. Stabler. The barcelona network, 2016.

[234] Ross M Thompson and CR Townsend. Impacts on stream food webs of native and exotic forest: An intercontinental comparison. *Ecology*, 84(1):145–161, 2003.

[235] Neo D Martinez. Artifacts or attributes? Effects of resolution on the Little Rock Lake food web. *Ecological monographs*, 61(4):367–392, 1991.

[236] Daniel J Bumbarger, Metta Riebesell, Christian Rödelsperger, and Ralf J Sommer. System-wide rewiring underlies behavioral differences in predatory and bacterial-feeding nematodes. *Cell*, 152(1):109–119, 2013.

[237] Neli Blagus, Lovro Šubelj, and Marko Bajec. Self-similar scaling of density in complex real-world networks. *Physica A: Statistical Mechanics and its Applications*, 391(8):2794–2802, 2012.

[238] Samin Aref, David Friggens, and Shaun Hendy. Analysing scientific collaborations of New Zealand institutions using Scopus bibliometric data. In *Proceedings of the Australasian Computer Science Week Multiconference*, pages 1–10, 2018.

[239] B. Jones. Computational geometry database, 2002.

[240] James Coleman, Elihu Katz, and Herbert Menzel. The diffusion of an innovation among physicians. *Sociometry*, 20(4):253–270, 1957.

[241] Vladimir Batagelj and Andrej Mrvar. Some analyses of Erdos collaboration graph. *Social Networks*, 22(2):173–186, 2000.

[242] Benedek Rozemberczki, Ryan Davies, Rik Sarkar, and Charles Sutton. GEMSEC: Graph embedding with self clustering. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2019*, pages 65–72. ACM, 2019.

[243] Ara Cho, Junha Shin, Sohyun Hwang, Chanyoung Kim, Hongseok Shim, Hyojin Kim, Hanhae Kim, and Insuk Lee. WormNet v3: A network-assisted hypothesis-generating server for Caenorhabditis elegans. *Nucleic acids research*, 42(W1):W76–W82, 2014.

[244] Robert E Ulanowicz and Donald L DeAngelis. Network analysis of trophic dynamics in south florida ecosystems. *FY97: The Florida Bay Ecosystem*, pages 20688–20038, 1998.

[245] Carlos J Melián and Jordi Bascompte. Food web cohesion. *Ecology*, 85(2):352–358, 2004.

[246] Laurent Dinh, Razvan Pascanu, Samy Bengio, and Yoshua Bengio. Sharp minima can generalize for deep nets. In *International Conference on Machine Learning*, pages 1019–1028. PMLR, 2017.

[247] C Daniel Freeman and Joan Bruna. Topology and geometry of half-rectified network optimization. In *International Conference on Learning Representations*, 2017.

[248] Yaim Cooper. The loss landscape of overparameterized neural networks. *arXiv preprint arXiv:1804.10200*, 2018.

[249] Berfin Simsek, François Ged, Arthur Jacot, Francesco Spadaro, Clément Hongler, Wulfram Gerstner, and Johanni Brea. Geometry of the loss landscape in overparameterized neural networks: Symmetries and invariances. In *International Conference on Machine Learning*, pages 9722–9732. PMLR, 2021.

[250] Hidenori Tanaka, Daniel Kunin, Daniel L Yamins, and Surya Ganguli. Pruning neural networks without any data by iteratively conserving synaptic flow. *Advances in neural information processing systems*, 33:6377–6389, 2020.

[251] Daniel Kunin, Javier Sagastuy-Brena, Surya Ganguli, Daniel LK Yamins, and Hidenori Tanaka. Neural mechanics: Symmetry and broken conservation laws in deep learning dynamics. *arXiv preprint arXiv:2012.04728*, 2020.

[252] Simon S Du, Wei Hu, and Jason D Lee. Algorithmic regularization in learning deep homogeneous models: Layers are automatically balanced. *Advances in neural information processing systems*, 31, 2018.

[253] Tengyuan Liang, Tomaso Poggio, Alexander Rakhlin, and James Stokes. Fisher-rao metric, geometry, and complexity of neural networks. In *The 22nd international conference on artificial intelligence and statistics*, pages 888–896. PMLR, 2019.

[254] Shun-ichi Amari, Hyeyoung Park, and Tomoko Ozeki. Geometrical singularities in the neuromanifold of multilayer perceptrons. *Advances in neural information processing systems*, 14, 2001.

[255] Justin Sirignano and Konstantinos Spiliopoulos. Mean field analysis of neural networks: A law of large numbers. *SIAM Journal on Applied Mathematics*, 80(2):725–752, 2020.

[256] Song Mei, Andrea Montanari, and Phan-Minh Nguyen. A mean field view of the landscape of two-layer neural networks. *Proceedings of the National Academy of Sciences*, 115(33):E7665–E7671, 2018.

[257] Grant Rotskoff and Eric Vanden-Eijnden. Trainability and accuracy of artificial neural networks: An interacting particle system approach. *Communications on Pure and Applied Mathematics*, 75(9):1889–1935, 2022.

[258] Kaifeng Lyu, Zhiyuan Li, Runzhe Wang, and Sanjeev Arora. Gradient descent on two-layer nets: Margin maximization and simplicity bias. *Advances in Neural Information Processing Systems*, 34:12978–12991, 2021.

[259] Etienne Boursier, Loucas Pillaud-Vivien, and Nicolas Flammarion. Gradient flow dynamics of shallow relu networks for square loss and orthogonal inputs. *Advances in Neural Information Processing Systems*, 35:20105–20118, 2022.

[260] Lenaic Chizat and Francis Bach. Implicit bias of gradient descent for wide two-layer neural networks trained with the logistic loss. In *Conference on learning theory*, pages 1305–1338. PMLR, 2020.

[261] Itay Safran, Gal Vardi, and Jason D Lee. On the effective number of linear regions in shallow univariate relu networks: Convergence guarantees and implicit bias. *Advances in Neural Information Processing Systems*, 35:32667–32679, 2022.

[262] Simon Eberle, Arnulf Jentzen, Adrian Riekert, and Georg S Weiss. Existence, uniqueness, and convergence rates for gradient flows in the training of artificial neural networks with relu activation. *arXiv preprint arXiv:2108.08106*, 2021.

[263] Bo Zhao, Iordan Ganev, Robin Walters, Rose Yu, and Nima Dehmamy. Symmetries, flat minima, and the conserved quantities of gradient flow. In *International Conference on Learning Representations*, 2023.

[264] Marco Nurisso, Pierrick Leroy, and Francesco Vaccarino. Topological obstruction to the training of shallow relu neural networks. *Advances in Neural Information Processing Systems*, 37:35358–35387, 2024.

[265] David Barrett and Benoit Dherin. Implicit gradient regularization. In *International Conference on Learning Representations*, 2020.

[266] Samuel L Smith, Benoit Dherin, David Barrett, and Soham De. On the origin of implicit regularization in stochastic gradient descent. In *International Conference on Learning Representations*, 2021.

[267] Quynh Nguyen. On connected sublevel sets in deep learning. In *International conference on machine learning*, pages 4790–4799. PMLR, 2019.

[268] Felix Draxler, Kambis Veschgini, Manfred Salmhofer, and Fred Hamprecht. Essentially no barriers in neural network energy landscape. In *International conference on machine learning*, pages 1309–1318. PMLR, 2018.

[269] Timur Garipov, Pavel Izmailov, Dmitrii Podoprikhin, Dmitry P Vetrov, and Andrew G Wilson. Loss surfaces, mode connectivity, and fast ensembling of dnns. *Advances in neural information processing systems*, 31, 2018.

[270] Itay Safran and Ohad Shamir. Spurious local minima are common in two-layer relu neural networks. In *International conference on machine learning*, pages 4433–4441. PMLR, 2018.

[271] Luca Venturi, Afonso S Bandeira, and Joan Bruna. Spurious valleys in one-hidden-layer neural network optimization landscapes. *Journal of Machine Learning Research*, 20(133):1–34, 2019.

[272] Marco Gori, Alessandro Betti, and Stefano Melacci. *Machine Learning: A constraint-based approach*. Elsevier, 2023.

[273] Wen-Liang Hwang and Shih-Shuo Tung. Analysis of function approximation and stability of general dnns in directed acyclic graphs using un-rectifying analysis. *Electronics*, 12(18):3858, 2023.

[274] Jonathan Frankle and Michael Carbin. The lottery ticket hypothesis: Finding sparse, trainable neural networks. In *International Conference on Learning Representations*, 2019.

[275] Torsten Hoefler, Dan Alistarh, Tal Ben-Nun, Nikoli Dryden, and Alexandra Peste. Sparsity in deep learning: Pruning and growth for efficient inference and training in neural networks. *Journal of Machine Learning Research*, 22(241):1–124, 2021.

[276] Gongfan Fang, Xinyin Ma, Mingli Song, Michael Bi Mi, and Xinchao Wang. Depgraph: Towards any structural pruning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16091–16101, 2023.

[277] Sumio Watanabe. *Algebraic geometry and statistical learning theory*, volume 25. Cambridge university press, 2009.

[278] Giovanni Luca Marchetti, Vahid Shahverdi, Stefano Mereta, Matthew Trager, and Kathlén Kohn. Position: Algebra unveils deep learning-an invitation to neuroalgebraic geometry. In *Forty-second International Conference on Machine Learning Position Paper Track*, 2025.

[279] Shun-ichi Amari and Tomoko Ozeki. Differential and algebraic geometry of multilayer perceptrons. *IEICE transactions on fundamentals of electronics, communications and computer sciences*, 84(1):31–38, 2001.

[280] Shun-ichi Amari, Hyeyoung Park, and Tomoko Ozeki. Singularities affect dynamics of learning in neuromanifolds. *Neural computation*, 18(5):1007–1065, 2006.

[281] Francis H Clarke, Yuri S Ledyaev, Ronald J Stern, and Peter R Wolenski. *Nonsmooth analysis and control theory*, volume 178. Springer Science & Business Media, 2008.

[282] Behnam Neyshabur, Russ R Salakhutdinov, and Nati Srebro. Path-sgd: Path-normalized optimization in deep neural networks. *Advances in neural information processing systems*, 28, 2015.

[283] William Wolberg, Olvi Mangasarian, Nick Street, and W Street. Breast cancer wisconsin (diagnostic). *UCI Machine Learning Repository*, 10:C5DW2B, 1995.

[284] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015.

[285] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256. JMLR Workshop and Conference Proceedings, 2010.

[286] Santiago López de Medrano. *Topology and Geometry of Intersections of Ellipsoids in R^ n*, volume 361. Springer Nature, 2023.

[287] L. R. Ford and D. R. Fulkerson. *Flows in Networks*. Princeton University Press, 1962.

[288] Ovidiu Calin. Neuromanifolds. In *Deep Learning Architectures: A Mathematical Approach*, pages 465–504. Springer, 2020.

[289] Kathlén Kohn. The geometry of the neuromanifold. *Collections*, 57(06), 2024.

[290] Tommaso Boccato, Matteo Ferrante, Andrea Duggento, and Nicola Toschi. 4ward: A relayering strategy for efficient training of arbitrarily complex directed acyclic graphs. *Neurocomputing*, 568:127058, 2024.

[291] Hassan Salehi Fathabadi and Mehdi Ghiyasvand. A new algorithm for solving the feasibility problem of a network flow. *Applied Mathematics and Computation*, 192(2):429–438, 2007.

[292] Alan J Hoffman. Some recent applications of the theory of linear inequalities to extremal combinatorial analysis. *New York, NY*, pages 113–117, 1958.

[293] Nathan W Henry, Giovanni Luca Marchetti, and Kathlén Kohn. Geometry of lightning self-attention: Identifiability and dimension. In *The Thirteenth International Conference on Learning Representations*, 2025.

[294] Vahid Shahverdi, Giovanni Luca Marchetti, and Kathlén Kohn. Learning on a razor's edge: the singularity bias of polynomial neural networks. *arXiv preprint arXiv:2505.11846*, 2025.

[295] Matthew Trager, Kathlen Kohn, and Joan Bruna. Pure and spurious critical points: a geometric study of linear networks. In *international conference on learning representations*, 2020.

[296] Vahid Shahverdi. Algebraic complexity and neurovariety of linear convolutional networks. *Acta Universitatis Sapientiae, Mathematica*, 17(1):2, 2025.

[297] Yossi Arjevani, Joan Bruna, Joe Kileel, Elzbieta Polak, and Matthew Trager. Geometry and optimization of shallow polynomial networks. *arXiv preprint arXiv:2501.06074*, 2025.

[298] Sergio Blanes, Fernando Casas, Jose-Angel Oteo, and José Ros. The magnus expansion and some of its applications. *Physics reports*, 470(5-6):151–238, 2009.

[299] Yun-Bin Zhao. An approximation theory of matrix rank minimization and its application to quadratic equations. *Linear Algebra and its Applications*, 437(1):77–93, 2012.

[300] William Wolberg, Olvi Mangasarian, Nick Street, and W. Street. Breast cancer wisconsin (diagnostic). UCI Machine Learning Repository, 1993. DOI: https://doi.org/10.24432/C5DW2B.