



Topological Signal Processing Over Generalized Cell Complexes

Stefania Sardellitti , Senior Member, IEEE and Sergio Barbarossa , Life Fellow, IEEE

Abstract—Topological Signal Processing (TSP) over simplicial complexes is a framework that has been recently proposed, as a generalization of graph signal processing (GSP), to extend GSP to analyzing signals defined over sets of any order (i.e., not only vertices of a graph) and to capture *multiway* relations of any order among the data. However, simplicial complexes are required to satisfy the so-called *inclusion property*, according to which, if a set belongs to the complex, then all its subsets must also belong to the complex. In some applications, this is a severe limitation. To overcome this limit, in this paper we extend TSP to deal with signals defined over cell complexes and we also generalize the concept of cell complexes to include hollow cells. We show that, even if the algebraic formulation does not change significantly, the extension to the generalized cell complexes considerably broadens the number of applications. Most important, the new representation provides a much better trade-off between the complexity of the representation and its accuracy. In addition, we propose a method to infer the structure of the cell complex from data and we propose distributed filtering strategies, including a method to retrieve the sparsest representation of the harmonic component. We quantify the advantages of using cell complexes instead of simplicial complexes, in terms of the complexity/accuracy trade-off, for different applications such as image segmentation and recovering of real flows measured on data traffic and transportation networks.

Index Terms—Topological signal processing, algebraic topology, cell complexes, simplicial complexes, FIR filters, topology inference.

I. INTRODUCTION

IN the last years a fast-growing interest emerged in signal processing and machine learning communities to develop models and tools for analyzing data defined over topological spaces [2], i.e. over non-metric domains composed of a set of points along with a set of neighborhood relations. A relevant

Manuscript received 4 October 2023; accepted 9 December 2023. Date of publication 12 January 2024; date of current version 30 January 2024. This work was supported in part by the H2020 EU/Taiwan Project 5G CONNI under Grant AMD-861459-3, in part by the MIUR under the PRIN Liquid-Edge Contract, and in part by the Huawei Technology France SASU under Agreement TC20220919044. An earlier version of this paper was presented in part at the 2021 IEEE Asilomar Conference [DOI: 10.1109/IEEECONF53345.2021.9723256]. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Gabriel Falcao. (Corresponding author: Stefania Sardellitti.)

The authors are with the Department of Information Engineering, Electronics, and Telecommunications, Sapienza University of Rome, 00184 Rome, Italy (e-mail: stefania.sardellitti@uniroma1.it; sergio.barbarossa@uniroma1.it).

This article has supplementary downloadable material available at <https://doi.org/10.1109/TSP.2023.3347069>, provided by the authors.

Digital Object Identifier 10.1109/TSP.2023.3347069

example is Graph Signal Processing (GSP) [3], a recent research field arising at the intersection between signal processing and graph theory, which transposes classical signal processing operations, such as, for example, filtering, sampling, estimation, to signals defined over the vertices of a graph. Graphs are a simple example of a topological space able to capture *pairwise* (or dyadic) relations between data associated with their vertices by encoding these relations through the presence of edges. However, in many complex systems such as gene regulatory networks, biological, social and brain networks, the richness of the interactions among the constitutive elements cannot be reduced to simple dyadic relationships [4]. To overcome the limitations of graph-based representations, it is necessary to go beyond graphs by incorporating *multiway* (or polyadic) relational descriptors. This generalization is possible resorting to *hypergraphs*, i.e. very general structures composed of a set \mathcal{V} of elements along with an ensemble \mathcal{S} of subsets of \mathcal{V} of any order [5]. However, the very generality of hypergraphs has an inherent complexity that can limit their applicability. To derive numerically efficient techniques able to handle signals defined over higher-order structures, it is useful to endow the domain with a hierarchical structure that enables processing at different levels. Examples of hypergraphs possessing a hierarchical structure, besides graphs, are *simplicial complexes* [6] and *cell complexes* [7], [8]. Very recently, combinatorial complexes have also been proposed as a very general hierarchical structure [9]. Simplicial complexes handle relations of any order, but are prone to respect the so called *inclusion property*, stating that if a set belongs to the space, all its subsets must belong to the space as well. In many applications, this constraint represents an unnecessary and unjustified limitation. Conversely, combinatorial complexes do not need to respect the inclusion property, but they lack of an algebraic structure that is useful in deriving signal processing tools. For all these reasons, in this paper we resort to cell complexes, which are quite general, lead to a natural hierarchical structure and can be formally described through an algebraic framework that provides the formal tools to *identify global properties, or invariants, of the signal domain by working with local operators*.

When analyzing data defined over a topological space, two main approaches are possible. The first class of methods is *Topological Data Analysis (TDA)* [2], where the data are mapped into a topological space and the information is extracted in terms of properties of the space. Persistent homologies represent one of the fundamental tools in this framework [10]. Methods based on simplicial complexes have been already

applied in many fields, such as statistical ranking [11], control systems [12], tumor progression analysis [13], and brain networks [14]. The two books [15], [16] focus on topological and geometric methods to analyze signals and images. The second class of methods is *Topological Signal Processing (TSP)*, where the observation is composed of a set of signals (data, features) defined over the subsets of a topological space and the information is extracted by processing these signals using *tools tailored to the properties of the space* [17], [18], [19]. Our goal in this paper is to extend the TSP framework developed in [17] to signals defined over a broad class of cell complexes. In TSP, the signals are real values (vectors) associated to the elements of the topological space. For example, there are applications to edge signals, i.e. signals associated to the edge of a network, in gene regulatory networks [20], neuronal signals [21] flowing between different areas of the brain, traffic flows in communication and transportation networks [22] and cross-disciplinary citations [23]. The analysis of edge signals has also been addressed in previous works as [24], [25], [26]. In [24] the authors introduce a class of filters based on the edge-Laplacian, while in [25] a semi-supervised learning method for divergence-free and curl-free edge flows was proposed. In [26] FIR filters for signals defined over simplices are considered.

A particularly relevant application of TSP is machine learning, and more specifically deep neural networks (DNNs). A message passing algorithm operating over simplicial complexes was proposed in [27]. Examples of DNNs operating over signals defined over simplicial complexes have been considered in [28], [29], [30]. An introduction to the processing of signals over cell complexes has been presented in [31], where it was shown that convolutional filters can serve as building blocks in neural networks defined over cell complexes. Recently, in [32], [33] novel attention-based neural networks have been defined over cell complexes. At a more fundamental level, one of the key unresolved issues in DNN is explainability. Some very interesting recent attempts have been proposed in [34], where the authors show that the layers of a large class of DNN can be interpreted as max-affine spline operators (MASOs) that partition the input space and then apply affine mappings to their input to produce the output. This partition of the input space can be interpreted as a cell complex. In such an application, the topological approach can be useful to identify invariants of the space that can help in extracting relevant information from the input data.

Typically, when embedding a cell complex into a real domain, the cells are assumed to be homeomorphic¹ to Euclidean balls. In this paper, besides extending TSP to conventional cell complexes, we will also introduce a new generalized class of complexes composed of disjoint polygons that may contain holes. A motivating example for introducing this new class of complexes is reported in Fig. 1. This image has been obtained as follows. We start from a gray-scale image, defined over a rectangular grid within a rectangle S . Each pixel is located in a position (x_i, y_i) and is the center of a square within which



Fig. 1. Cell complex induced by an image-dependent space partition.

the image assumes a constant value $f(x_i, y_i)$. The ensemble of all the squares represents a partition of the overall domain S . In this way, we can think of the image as a 2D (piecewise) function $f(x, y)$, defined over the continuous domain S . Then, we quantize the image by selecting a discrete number of $H + 1$ levels f_0, f_1, \dots, f_H , and we identify the following domains:

$$\mathcal{D}_i \equiv \{(x, y) : f_i < f(x, y) \leq f_{i+1}\}, i = 0, \dots, H - 1. \quad (1)$$

It is easy to check that the ensemble of all these domains represents a partition of S . By construction, each domain \mathcal{D}_i is composed of polygons, possibly disconnected and having holes in their interior. These polygons are examples of generalized cells that, differently from the conventional case, are not homeomorphic to a ball in 2D domain. In computational topology, some algorithms have been developed to find a 2D/3D space partition using a collection of geometric objects that can possibly contain holes, see e.g. [35], [36], but no TSP tools have been introduced to study signals defined over these domains.

This paper is an extension of our preliminary work [1], and it contains the following contributions:

- a) we introduce a new topological structure that generalizes cell complexes to include polygons with holes. The structure is especially suitable in image processing;
- b) we show how the proposed complex yields a better sparsity/accuracy trade-off with respect to graphs or simplicial complexes;
- c) we propose a new method to extract a compact representations of the invariants characterizing the signal domain;
- d) we propose a method to infer the cell complex structure from data;
- e) we introduce a distributed filtering of signals defined over cell complexes that extends the approaches proposed in [26], [37].

¹A homeomorphism is a mapping between spaces that preserves all their topological properties. Two topological spaces are homeomorphic if there is a continuous and invertible mapping between them.

The paper is organized as follows. In Section II we recall the basic topological tools, distinguishing between abstract and geometric cell complexes; we also propose a generalization of cell complexes to the case where the (2D) cells may contain holes. In Section III we introduce their algebraic representation, while in Section IV we show how to represent signals defined over cell complexes. In Section V we propose a method to infer the cell complex structure from data and we show how the cell complex model yields a better sparsity/accuracy trade-off with respect to graph- or simplicial-based methods. In Section VI we show how to extract compact representations of the invariants (homologies) associated to the cell structure. In Section VII, we introduce the design of FIR filters for the solenoidal and irrotational signal components. Finally, in Section VIII we draw some conclusions.

II. INTRODUCTION TO ALGEBRAIC TOPOLOGICAL TOOLS

In this section, we recall the basic tools to represent cell complexes, distinguishing between abstract and geometric cell complexes.

A. Abstract Cell Complexes

As illustrated in [38], the notion of cell complexes historically emerged with the rise of topology and is one of the most dynamic notions in the mathematical literature. In this work we focus on complexes composed of a finite number of cells. An abstract cell complex (ACC) is a finite partially ordered set, or *poset*, equipped with a dimension function (graded poset [39]), whose fundamental elements are called cells. A more formal definition of an ACC is the following [38]:

Definition 1: An abstract cell complex $\mathcal{C} = \{\mathcal{S}, \prec_b, \dim\}$ is a set \mathcal{S} of abstract elements (cells) provided with a binary relation \prec_b , called the bounding (or incidence) relation, and with a dimension function, denoted by $\dim(x)$, that assigns to each $x \in \mathcal{S}$ a non-negative integer $[x]$, satisfying the following two axioms:

- A1) if $x \prec_b y$ and $y \prec_b z$, then $x \prec_b z$ follows (transitivity);
- A2) if $x \prec_b y$, then $\dim(x) < \dim(y)$ (monotonicity).

If $\dim(x) = n$, then n is the dimension (or order) of x and x is called an n -cell; 0-cells are named vertices. We denote a cell x of order n with x^n . In words, if $x, y \in \mathcal{S}$ and $x \prec_b y$, we say that x bounds y . If $x \prec_b y$ and $\dim(x) = \dim(y) - 1$ then x is a face of y and y is a co-face of x . We assume that every 1-cell of \mathcal{C} is incident with two 0-cells of \mathcal{C} and distinct 1-cells are not incident to the same pair of 0-cells. Given an n -dimensional cell x^n , we define its boundary ∂x^n as the set of all cells of dimension less than n that bound x^n . An ACC is a k -dimensional ACC, or simply a k -complex, if the dimensions of all its cells are less than or equal to k . We denote by $\bar{x}^n = x^n \cup \partial x^n$ the closed cell including its boundary. Two cells x, y are incident if $x \prec_b y$ or $y \prec_b x$. To associate a *topology* with an ACC, we need to define neighboring relations among the cells of the complex [40], [41]. This can be done by introducing the concept of *open* and *closed* subsets [38]. A subset U of $\mathcal{C} = \{\mathcal{S}, \prec_b, \dim\}$ is called *open* (or upper set) of \mathcal{C} iff, for every element x of U , all elements y of \mathcal{C}

for which $x \prec_b y$ are also in U . A subset U is *closed* (or down set) iff, for every element x of U , all elements y of \mathcal{C} for which $y \prec_b x$ are also in U .

An ACC equipped with the above neighboring relations is a topological space [41]. In signal processing, it is also useful to associate an algebraic structure with the topological space. To do that, we need to introduce an algebraic operator, called the *boundary operator*, which maps each cell to its boundary. This mapping is made possible if we assume that all cells belonging to the complex are *closed*: A cell x of order n is closed if all its bounding faces belong to the complex. The boundary ∂x is the sequence of all cells of order $n - 1$ that bound x . The boundary operator identifies each cell uniquely and makes possible the association of an algebraic representation with an abstract cell complex.

As an example, let us consider a simple ACC composed by 4 cells of order 0, 4 cells of order 1 and 1 cell of order 2. Denoting the i -th cell of order k as c_i^k , the set \mathcal{S} of abstract elements is defined by $\mathcal{S} = \{\{c_i^0\}_{i=1}^4, \{c_i^1\}_{i=1}^4, c_1^2\}$. The neighboring relations among the elements of the set \mathcal{S} are given by the four open sets $U_1 = \{c_1^0, c_1^1, c_4^1, c_1^2\}$, $U_2 = \{c_2^0, c_4^1, c_3^1, c_1^2\}$, $U_3 = \{c_3^0, c_1^1, c_2^1, c_1^2\}$, $U_4 = \{c_4^0, c_2^1, c_3^1, c_1^2\}$. These open sets fully describe the cell complex \mathcal{C} . Note that the boundary ∂c_1^2 of the second order cell c_1^2 is defined by the closed sequence of incident edges $[c_1^1, c_2^1, c_3^1, c_4^1]$, where all edges belong to the complex. Since the cell c_1^2 is closed, it can be uniquely identified through its bounding edges.

Note that the notion of abstract cell complex differs and it is much more general than that of a simplicial complex because it is not required to respect the inclusion property.

B. Geometric Cell Complexes

Differently from ACCs, a *geometric cell complex* (GCC) is a collection of finite-dimensional cells embedded in a Euclidean space. A geometric cell complex is a topological space built by sequentially attaching cells of increasing dimension along their boundaries. In the following, we first introduce regular geometric cell complexes whose cells are homeomorphic to Euclidean balls. Then, we will define more general cell complexes whose cells are not constrained to be homeomorphic to Euclidean balls but may contain holes.

Regular geometric cell complexes (RGCCs). The formal definition describing how the cells may be glued together in an RGCC, is the following [42]:

Definition 2: A regular cell complex is a topological space X with a partition $\{X_\sigma\}_{\sigma \in P_X}$ of subspaces X_σ of X , called cells, where P_X is a partially ordered set, satisfying the following conditions:

- 1) every open k -order cell X_σ is homeomorphic to the interior of a k -dimensional Euclidean ball;
- 2) denoting by ∂X_σ the boundary of X_σ , every closed k -order cell can be written as $\bar{X}_\sigma = X_\sigma \cup \partial X_\sigma$, and it is homeomorphic to a closed ball in a k -dimensional Euclidean space; ∂X_σ is then homeomorphic to the closure of the ball associated to X_σ ;
- 3) For all α, σ , $\bar{X}_\alpha \cap X_\sigma \neq \emptyset$ iff $X_\sigma \subseteq \bar{X}_\alpha$.

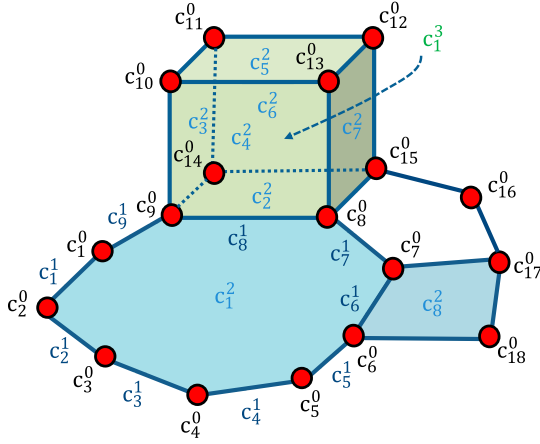


Fig. 2. A RGCC of order 3.

Since any two open cells are disjoint by definition, condition 3) implies that two closed cells can only intersect over their boundaries. Intuitively speaking, a regular cell complex allows us to represent a Euclidean space as a poset, i.e. a hierarchical structure that will turn out to be very useful in the analysis of signals defined over the structure. Every cell c_i has associated a number, called dimension. The n -skeleton X^n of order n of X is defined as $X^n = \bigcup_{\dim(c_i) \leq n} c_i$. The boundary of a cell c of dimension n is the set of all cells of dimension $n - 1$ that bound c . As with ACCs, we assume that every 1-cell of \mathcal{C} is incident with two 0-cells of \mathcal{C} and distinct 1-cells are not incident to the same pair of 0-cells.

As an example, a closed 1-cell (edge) can be built from two 0-cells (points c_1^0, c_2^0) and an interval I (1-dimensional ball) by glueing one endpoint of I to c_1^0 and the other to c_2^0 . A k -cell may be represented by an ordered set of vertices comprising a k -polytope². The dimension or order of a cell complex is the largest dimension of any of its cells. A graph is a particular case of a cell complex of order 1, containing only cells of order 0 (nodes) and 1 (edges). Simplicial complexes are special cases of cell complexes for which any k -cell is composed of exactly $k + 1$ -vertices. In Fig. 2 we show an example of a 3-dimensional GCC defined by the set of elements $\mathcal{S} = \{\{c_i^0\}_{i=1}^{18}, \{c_i^1\}_{i=1}^{25}, \{c_i^2\}_{i=1}^8, c_1^3\}$. The illustrated structure is not a simplicial complex, because the presence of the nonagon (cell c_1^1) and the closed cube (cell c_1^3) does not imply the presence of all their subsets. Notice also that the 2-order closed cell c_1^2 contains more 0-order cells (vertices) than the 3-order closed cell c_1^3 . This is a situation that cannot occur in a simplicial complex.

Planar hollow cell complexes. The previous definition of geometric cell complex does not encompass the complex depicted in Fig. 1, where some polygons contain holes. In this section, we generalize the previous definition of complex to be able to handle signals as in Fig. 1. More specifically, we introduce a new class of complexes, that we name *planar*

hollow cell complex, whose distinguishing feature is that its cells are not constrained to be homeomorphic to a 2D ball, but can contain a finite number of holes in their interior. More specifically, we use the following definition:

Definition 3: Given a finite region \mathcal{D} of \mathbb{R}^2 , a Planar Hollow Cell Complex (PHCC) is a partition of \mathcal{D} with a finite number of cells $\{X_\sigma\}_{\sigma \in P_X}$ satisfying the following conditions:

- 1) The 1-skeleton associated with X is a planar graph³;
- 2) Every k -cell X_σ , with $k = 0, 1$, is homeomorphic to the open k -ball in \mathbb{R}^k ;
- 3) Every hollow 2-cell X_σ is a path-connected region that is homeomorphic to a 2-ball in \mathbb{R}^2 that may include a finite number n_h of 2D balls, with $n_h \geq 0$.
- 4) The boundary of a hollow 2-cell, including n_h holes in its interior, is composed of $n_h + 1$ disconnected closed paths, each one belonging to a different graph component. Every cell is a path-connected space enclosed between the outer and the n_h inner boundaries of the n_h holes; each hole may be filled with other cells or it may be empty.
- 5) A closed cell \bar{X}_α is defined as $\bar{X}_\alpha = X_\alpha \cup \partial X_\alpha$. For cells X_α of order $k = 0, 1$ the boundary ∂X_α is homeomorphic to the closure of the ball associated with X_α . For all α, σ , $\bar{X}_\alpha \cap X_\sigma \neq \emptyset$ iff $X_\sigma \subseteq \bar{X}_\alpha$.

As a pictorial example, Fig. 3(a) shows a PHCC composed of a hollow cell with one hole (the blue cell c_1^2), filled with two triangular cells with no holes (the green cells c_2^2 and c_3^2).

III. ALGEBRAIC REPRESENTATION OF CELL COMPLEXES

The structure of a cell complex is captured by the neighborhood relations among its cells. As with graphs, it is useful to introduce the orientation of the cells first. The orientation of a cell can be derived by generalizing the concept of orientation of a simplex. Every simplex can only have two orientations, depending on the permutations of its elements. Defining the transposition as the permutation of two elements, two orientations are equivalent if each of them can be recovered from the other through an even number of transpositions [43].

To define the orientation of k -cells, we may apply first a simplicial decomposition [8], which consists in subdividing the cell into a set of internal k -simplices, so that: i) two simplices share exactly one $(k - 1)$ -simplicial boundary element, which is not the boundary of any other k -cell in the complex; and ii) two k -simplices induce an opposite orientation on the shared $(k - 1)$ -boundary. Hence, by orienting a single internal simplex, the orientation propagates on the entire cell.

Abstract Cell complexes and GCCs. An oriented k -cell of an ACC or a GCC may be represented through its bounding cells as $c^k = [c_1^{k-1}, \dots, c_M^{k-1}]$ where two consecutive $(k - 1)$ -cells, c_i^{k-1} and c_{i+1}^{k-1} share a common $(k - 2)$ -cell boundary. We use the notation $c_i^{k-1} \sim c_j^k$ to indicate that the orientation of c_i^{k-1} is coherent with that of c_j^k and $c_i^{k-1} \approx c_j^k$ to indicate that their orientations are opposite. Given an orientation, there are two ways in which two cells can be considered to be adjacent: lower and upper adjacent. Two k -order cells are lower adjacent

²A k -polytope is a geometric object with flat sides, i.e. with sides consisting of $(k - 1)$ -polytopes that may have $(k - 2)$ -polytopes in common.

³A planar graph is a graph that can be embedded in the plane in a such a way that its edges intersect only at their endpoints.

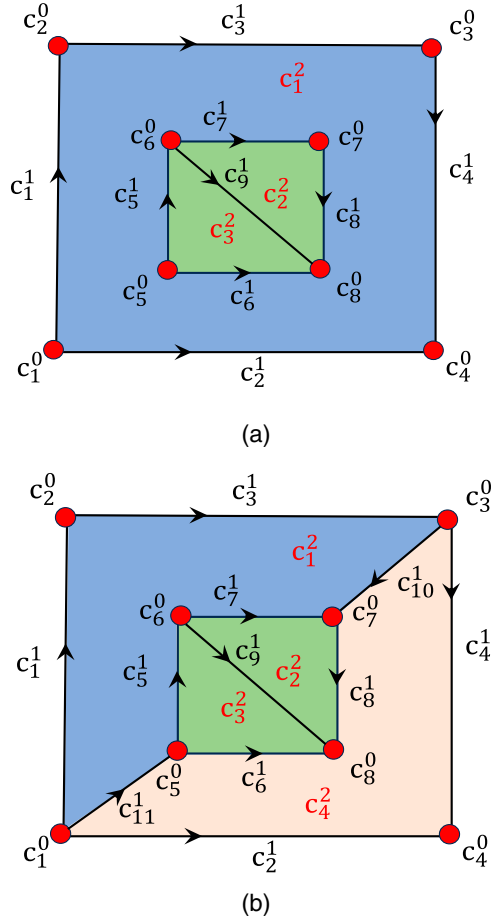


Fig. 3. (a) A planar hollow cell complex; (b) a planar polygonal cell complex.

if they share a common face of order $k - 1$ and upper adjacent if they are both faces of a cell of order $k + 1$. Given an orientation of the cell complex \mathcal{C} , the structure of a cell complex of order K is fully captured by the set of its incidences matrices \mathbf{B}_k with $k = 1, \dots, K$, also named boundaries matrices, whose entries establish which k -cells are incident to which $(k - 1)$ -cells and are defined as follows:

$$B_k(i, j) = \begin{cases} 0, & \text{if } c_i^{k-1} \not\prec_b c_j^k \\ 1, & \text{if } c_i^{k-1} \prec_b c_j^k \text{ and } c_i^{k-1} \sim c_j^k \\ -1, & \text{if } c_i^{k-1} \prec_b c_j^k \text{ and } c_i^{k-1} \approx c_j^k \end{cases} \quad (2)$$

Let us consider a cell complex of order two $\mathcal{C} = \{\mathcal{V}, \mathcal{E}, \mathcal{P}\}$ where \mathcal{V} , \mathcal{E} , \mathcal{P} denote the set of 0, 1 and 2-cells, i.e. vertices, edges and polygons, respectively. We denote their cardinality by $|\mathcal{V}| = N$, $|\mathcal{E}| = E$ and $|\mathcal{P}| = P$. Then, the two incidence matrices describing the connectivity of the complex are $\mathbf{B}_1 \in \mathbb{R}^{N \times E}$ and $\mathbf{B}_2 \in \mathbb{R}^{E \times P}$, where \mathbf{B}_2 can be written as

$$\mathbf{B}_2 = [\mathbf{B}_T, \mathbf{B}_Q, \dots, \mathbf{B}_{P_{max}}] \quad (3)$$

with \mathbf{B}_T , \mathbf{B}_Q and $\mathbf{B}_{P_{max}}$ describing the incidences between edges and, respectively, triangles, quadrilateral, up to polygons with P_{max} sides. To build the incidence matrices we need to find the edges bounding each polygon in the complex. Defining a cycle as a sequence of distinct and lower adjacent edges,

starting from and ending in the same node, we have to search for all chordless cycles. To do that, we first find all cycles of increasing length, from 3 up to a maximum value P_{max} , by using a numerical graph toolbox, for example, the matlab function all cycles. Then, we eliminate every cycle of length n that can be obtained as a linear combination of cycles of length smaller than n . The remaining cycles identify the boundary edges of each 2-cell and are then used to build the columns of the incidence matrices in (3).

An important property is that the boundary of a boundary is zero, i.e. it always holds $\mathbf{B}_k \mathbf{B}_{k+1} = \mathbf{0}$. To describe the structure of the K -cell complex we can consider the higher order combinatorial Laplacian matrices [44] given by

$$\begin{aligned} \mathbf{L}_0 &= \mathbf{B}_1 \mathbf{B}_1^T, \\ \mathbf{L}_k &= \mathbf{B}_k^T \mathbf{B}_k + \mathbf{B}_{k+1} \mathbf{B}_{k+1}^T \text{ for } k = 1, \dots, K - 1 \\ \mathbf{L}_K &= \mathbf{B}_K^T \mathbf{B}_K \end{aligned} \quad (4)$$

where $\mathbf{L}_{k,d} = \mathbf{B}_k^T \mathbf{B}_k$ and $\mathbf{L}_{k,u} = \mathbf{B}_{k+1} \mathbf{B}_{k+1}^T$ are the lower and upper Laplacians, expressing the lower and upper adjacency of the k -order cells, respectively. Note that \mathbf{L}_0 is the graph Laplacian. There are some interesting properties of the eigenvectors of higher order Laplacians [44], that it is useful to recall for our ensuing spectral analysis. Considering w.l.o.g. the first order Laplacian, i.e.

$$\mathbf{L}_1 = \mathbf{L}_{1,d} + \mathbf{L}_{1,u} = \mathbf{B}_1^T \mathbf{B}_1 + \mathbf{B}_2 \mathbf{B}_2^T, \quad (5)$$

it holds: i) the eigenvectors associated with the nonzero eigenvalues of $\mathbf{L}_{1,d} = \mathbf{B}_1^T \mathbf{B}_1$ are orthogonal to the eigenvectors associated with the nonzero eigenvalues of $\mathbf{L}_{1,u} = \mathbf{B}_2 \mathbf{B}_2^T$ and viceversa; ii) the eigenvectors associated with the nonzero eigenvalues λ_i^1 of \mathbf{L}_1 are either the eigenvectors of $\mathbf{L}_{1,d}$ or those of $\mathbf{L}_{1,u}$; and, finally, iii) the nonzero eigenvalues of \mathbf{L}_1 are either the eigenvalues of $\mathbf{L}_{1,d}$ or those of $\mathbf{L}_{1,u}$. This spectral structure of \mathbf{L}_1 induces an interesting decomposition of the whole space \mathbb{R}^E , the so-called Hodge decomposition [45], given by

$$\mathbb{R}^E \triangleq \text{img}(\mathbf{B}_1^T) \oplus \ker(\mathbf{L}_1) \oplus \text{img}(\mathbf{B}_2) \quad (6)$$

where the vectors in $\ker(\mathbf{L}_1)$ are also in $\ker(\mathbf{B}_1)$ and $\ker(\mathbf{B}_2^T)$.

Planar Hollow Cell Complexes. In this section, we generalize the standard cell complex approach to incorporate cells (polygons) containing one or more holes in their interior. What distinguishes hollow cells from the others is that they have one outer boundary and one or more inner boundaries, one for each hole. These boundaries are, by construction, cycles that are disconnected from each other. In this section, we show that also working with hollow cells, we can still extract the fundamental invariant properties of the space, such as the number of connected components or number of holes, from the algebraic representation given by the incidence matrices. We start with the illustrative example of the 2-order complex depicted in Fig. 3(a), and then we will provide the full generalization. The complex in Fig. 3(a) has an outer square cell with one inner hole, containing two triangular cells without holes. In this

case, the node-edge and edge-polygon incidence matrices are, respectively,

$$\mathbf{B}_1 = \begin{bmatrix} -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix},$$

$$\mathbf{B}_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & -1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{bmatrix}$$

where the first two columns of \mathbf{B}_2 identify the triangular cells c_2^2 and c_2^3 without holes, while the third column identifies the cell c_1^2 through its outer and inner bounding edges. Looking at \mathbf{B}_1 , we can see that it has a block structure, corresponding to the two disconnected graphs composing the 1-skeleton of the complex. The kernel of \mathbf{B}_1 has then a dimension 2, coinciding with the number of disconnected components. Furthermore, it is easy to verify that, as with standard cell complexes, $\mathbf{B}_1\mathbf{B}_2 = \mathbf{0}$. This implies, using standard algebra, see e.g., [45], that

$$\text{im}(\mathbf{B}_1^T\mathbf{B}_1 + \mathbf{B}_2\mathbf{B}_2^T) = \text{im}(\mathbf{B}_1^T) + \text{im}(\mathbf{B}_2).$$

In this example, the dimension of $\text{im}(\mathbf{B}_1^T)$ is $8 - 2 = 6$, while the dimension of $\text{im}(\mathbf{B}_2)$ is 3, since \mathbf{B}_2 is clearly full rank. Hence, the dimension of $\text{im}(\mathbf{L}_1)$ is $6 + 3 = 9$, which means that \mathbf{L}_1 is full rank. This implies that the number of holes, i.e. zero, is correctly identified with the dimension of the kernel of \mathbf{L}_1 . Furthermore, according to Euler's Theorem for planar graphs with n_c disconnected components [46], the number of 2-order faces in \mathcal{X} , excluding the unbounded external face, is given by $F = E - N + n_c$. Therefore, in our example we get $F = 3$. Moreover, we can check that, if we remove any face from the space, which means that we remove one column from the incidence matrix \mathbf{B}_2 , we create a hole in the space and indeed the dimension of $\text{ker}(\mathbf{L}_1)$ becomes equal to one, i.e. it coincides with the number of holes. In summary, the first two Betti numbers of the space sketched in Fig. 3(a) are then $\beta_0 = 2$ (number of connected components) and $\beta_1 = 0$ (number of holes). An interesting property to notice about filling this space with hollow cells is that if we had filled the same space with a structure whose 1-d skeleton was connected, as depicted in Fig. 3(b), we would have found $\beta_0 = 1$ and $\beta_1 = 0$. We prove now that the above results are valid in general, for any planar hollow cell complex.

Theorem 1: Assume that a 2-order cell complex $\mathcal{X} = \{\mathcal{V}, \mathcal{E}, \mathcal{F}\}$ defines a partition of a closed space $\mathcal{D} \subseteq \mathbb{R}^2$ through its cells c_i^k of order $k = 0, 1, 2$, with $\bigcup_{i,k} c_i^k = \mathcal{D}$. The 1-skeleton associated with the complex is composed by

the union of $n_c \geq 1$ disconnected graphs $\mathcal{G}_i = \{\mathcal{V}_i, \mathcal{E}_i\}$. Then, it holds:

- The cell complex \mathcal{X} admits an algebraic representation through the node-edge and edge-polygon incidence matrices \mathbf{B}_1 and \mathbf{B}_2 (given, respectively, in (40) and (41) of the supplementary material) and the following orthogonality condition holds

$$\mathbf{B}_1\mathbf{B}_2 = \mathbf{0}. \quad (7)$$

- The structure of the planar complex is described by the first-order Laplacian matrix $\mathbf{L}_1 = \mathbf{B}_1^T\mathbf{B}_1 + \mathbf{B}_2\mathbf{B}_2^T$, which implies that the Hodge decomposition still holds, so that the whole space \mathbb{R}^E is the direct sum of three orthogonal subspaces, the irrotational, solenoidal and harmonic subspaces, as in (6). The dimension of $\text{ker}(\mathbf{L}_1)$ counts the number of 1-dimensional holes in the complex, i.e. the regions of the plane that remain uncovered.

Proof: See Appendix A in the supplementary material. ■

IV. SIGNAL PROCESSING OVER CELL COMPLEXES

Some fundamental tools to analyze signals defined over simplicial complexes are provided in [8], [17], [18], [19]. Here we show that the framework can be easily extended to cell complexes. We focus w.l.o.g. on cell complexes of order 2. Given a cell complex $\mathcal{C} = \{\mathcal{V}, \mathcal{E}, \mathcal{P}\}$, the signal on vertices, edges and polygons are defined by the following maps: $s^0 : \mathcal{V} \rightarrow \mathbb{R}^N$, $s^1 : \mathcal{E} \rightarrow \mathbb{R}^E$, and $s^2 : \mathcal{P} \rightarrow \mathbb{R}^P$. Exploiting the Hodge decomposition in (6), we may always express a signal s^1 as the sum of three orthogonal components [45], i.e.

$$s^1 = \mathbf{B}_1^T s^0 + \mathbf{B}_2 s^2 + s_H^1. \quad (8)$$

In analogy to vector calculus terminology, the first component $s_{irr}^1 := \mathbf{B}_1^T s^0$ is called the irrotational component since, using the equality $\mathbf{B}_1\mathbf{B}_2 = \mathbf{0}$, it has zero curl, i.e. $\mathbf{B}_2^T s_{irr}^1 = \mathbf{0}$. The second term $s_{sol}^1 := \mathbf{B}_2 s^2$ is the solenoidal component, since its divergence defined as $\mathbf{B}_1 s_{sol}^1$ is zero. Finally, the component s_H^1 is the harmonic component of the signal since it belongs to $\text{ker}(\mathbf{L}_1)$, and it is a signal with zero curl and zero divergence. Note from (8) that the signals defined over cells of order k depend on the signals defined over cells of order $k - 1$ and $k + 1$. This interdependence between signals of different orders is what allows us to encode and capture the topological structure of the data.

A useful orthogonal basis to represent signals of various order, capturing the connectivity properties of the complex, is given by the eigenvectors of the corresponding higher order Laplacian. Then, generalizing graph spectral theory [3], we can introduce, as for simplicial complex [17], a notion of Cell complex Fourier Transform (CFT) for signals defined over cell complexes. Let us consider the eigen decomposition

$$\mathbf{L}_k = \mathbf{U}_k \mathbf{\Lambda}_k \mathbf{U}_k^T \quad (9)$$

where \mathbf{U}_k is the eigenvectors matrix and $\mathbf{\Lambda}_k$ is a diagonal matrix with entries the eigenvalues λ_i^k of \mathbf{L}_k , with $i = 1, \dots, E$.

Then, we define the k -order CFT as the projection of a k -order signal onto the eigenvectors of \mathbf{L}_k , i.e.

$$\hat{\mathbf{s}}^k \triangleq \mathbf{U}_k^T \mathbf{s}^k. \quad (10)$$

A signal \mathbf{s}^k can then be represented in terms of its CFT coefficients as $\mathbf{s}^k = \mathbf{U}_k \hat{\mathbf{s}}^k$. Note that the eigenvectors of the upper and lower Laplacians are the optimal orthogonal vectors minimizing the quadratic forms $\mathbf{x}^{1T} \mathbf{B}_1^T \mathbf{B}_1 \mathbf{x}^1$ and $\mathbf{x}^{1T} \mathbf{B}_2 \mathbf{B}_2^T \mathbf{x}^1$, respectively. These quadratic forms represent the squared norm of the divergence and the curl vectors of \mathbf{x}^1 , respectively. Then, the first eigenvectors of the lower and upper Laplacians enable capturing the signals smoothness in terms of divergence and curl variations.

In [17], using the signal decomposition in (8) we analyzed optimization strategies for the estimation of the signals \mathbf{s}^0 , \mathbf{s}^2 and \mathbf{s}_H^1 from noisy observations of edge signals defined over simplicial complexes and we provided conditions for their sampling and reconstruction. Those results can be extended to signals defined over cell complexes because the structure of \mathbf{B}_2 in (3) (and (41) in the supplementary material) has the same properties, thus generalizing the approach of [17]. Let us now investigate how the use of cell complexes yields a substantial performance improvement in terms of sparsity/accuracy trade-off with respect to simplicial structures. As an illustrative application, we focus on an image coding task by showing how planar cell complexes are capable of partitioning the image domain in cells encoding similarity among the intensity of adjacent pixels. We consider, as an example, a gray-scale version of the Mona Lisa painting, shown in Fig. 1. The image is defined over a rectangular grid by associating each pixel to a square. Then, we first derive a simplicial representation of the image by building a gradient-based Delaunay triangulation of its domain using 6100 triangles and 12 gray levels. Hence, we get a 2D-dimensional function $f(x, y)$ with constant intensity (gray level) over each triangle. We denote with $s_2(i)$ the signal associated to each triangle i , i.e. the image intensity observed over it. Then, we build a planar cell complex whose cells are identified according to (1) where f_i , $i = 0, \dots, H$, are the selected gray levels. The result, as shown in Fig. 1, is a planar hollow cell complex composed of $N_c = 969$ cells, where many polygons have holes. To each cell we associate a 2-order signal $s_2^{cc}(k)$, given by the intensity f_i of the signals on the corresponding domain, as illustrated in Fig. 1. Clearly, this complex is tuned to the image and is then better able to grasp more structured relationships among the image pixels. For example, a big part of the hair of Mona Lisa gives rise to a big polygonal cell associated to pixels having similar intensity. It is also useful to notice that it is quite likely that the adopted space partition yields hollow cells, like for example the cell c_1^2 that has a hole represented by cell c_2^2 , as indicated by the two arrows in Fig. 1.

It is now interesting to compare the two inferred simplicial and planar cell complexes, in terms of image representation. Specifically, using the second order Laplacian matrices $\mathbf{L}_2 = \mathbf{U}_2 \mathbf{\Lambda}_{sc} \mathbf{U}_2^T$ and $\mathbf{L}_2^{cc} = \mathbf{U}_2^{cc} \mathbf{\Lambda}_{cc} \mathbf{U}_2^{ccT}$, associated with the SC and CC complexes, respectively, we project the signals \mathbf{s}_2 and

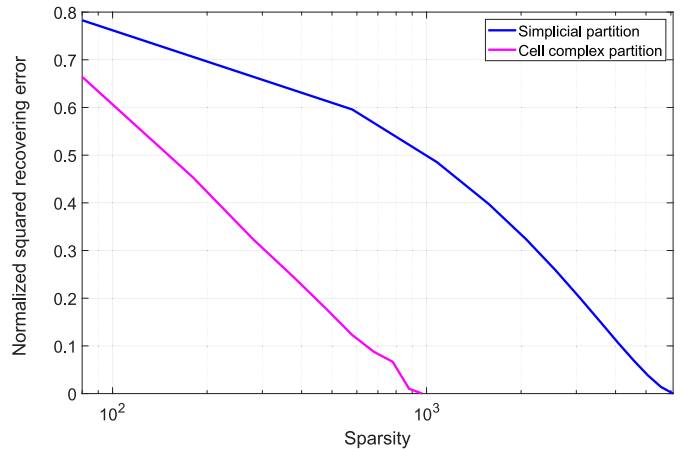


Fig. 4. Normalized squared error vs. the signal sparsity.

\mathbf{s}_2^{cc} over the eigenvectors bases \mathbf{U}_2 and \mathbf{U}_2^{cc} , thus obtaining the following simplicial and cell Fourier coefficients:

$$\hat{\mathbf{s}}_2 = \mathbf{U}_2^T \mathbf{s}_2, \quad \hat{\mathbf{s}}_2^{cc} = \mathbf{U}_2^{ccT} \mathbf{s}_2^{cc}. \quad (11)$$

Then, we build the matrices (bases) $\mathbf{U}_{2,\mathcal{K}}$ and $\mathbf{U}_{2,\mathcal{K}_c}^{cc}$ containing as columns the eigenvectors, respectively, with indexes in the sets \mathcal{K} and \mathcal{K}_c , associated with the highest K coefficients in (11). Then, we get the K -sparse signal representations $\hat{\mathbf{s}}_{2,\mathcal{K}} = \mathbf{U}_{2,\mathcal{K}}^T \mathbf{s}_2$ and $\hat{\mathbf{s}}_{2,\mathcal{K}_c}^{cc} = \mathbf{U}_{2,\mathcal{K}_c}^{ccT} \mathbf{s}_2^{cc}$. Finally, the images are reconstructed as

$$\bar{\mathbf{s}}_2 = \mathbf{U}_{2,\mathcal{K}} \hat{\mathbf{s}}_{2,\mathcal{K}}, \quad \bar{\mathbf{s}}_2^{cc} = \mathbf{U}_{2,\mathcal{K}_c}^{cc} \hat{\mathbf{s}}_{2,\mathcal{K}_c}^{cc}.$$

In Fig. 4, we plot the normalized mean squared error in the reconstructed image vs. its sparsity, i.e. the number K of bases (eigenvectors) used to represent the signal. It can be noticed the considerable gain in terms of accuracy/sparsity trade-off that a cell based representation is able to provide with respect to a simplicial complex partition. Intuitively speaking, the advantage comes from a better tuning of the cell complex to the image.

V. TOPOLOGY INFERENCE AND SPARSE REPRESENTATION

The goal of this section is to propose a method to infer the structure of a cell complex from data by generalizing the method proposed in [17] for the inference of simplicial complexes. The underlying idea is that the inclusion of polygons of order greater than three yields a richer dictionary that can be exploited to find a better trade-off between sparsity and representation accuracy with respect to the simplicial case.

The general method is inductive and it starts from a zero order set, i.e. an ensemble of points, and infers the structure of a first order set by exploiting the relations between the signals associated to the zero-order elements. This first step has been largely studied as the graph topology inference problem, see, e.g., [47], [48]. The successive steps are a generalization of the first step, every time considering an order up. So, in general, the goal of the n -th step is to infer the structure of an n -th order cell complex from signals defined over the $(n-1)$ -complex. We focus on the inference of the second order complex, but

the procedure can be generalized to higher order structures in a straightforward manner. Hence, in our case we start from an underlying graph, whose incidence matrix \mathbf{B}_1 is then known, and we want to infer the structure of \mathbf{B}_2 from the signals defined over the edges of the graph. We start from the observation of M snapshots of edge signals $\mathbf{x}^1(m)$, $m = 1, \dots, M$. As a first step, we check if it is really necessary to go beyond graphs and then introduce the matrix \mathbf{B}_2 to represent the observed data. Since the only signal components that may depend on \mathbf{B}_2 are the solenoidal and harmonic components, we project the observed signals $\mathbf{x}^1(m)$ for $m = 1, \dots, M$ onto the solenoidal and harmonic subspaces by computing

$$\mathbf{x}_{sH}^1(m) = (\mathbf{I} - \mathbf{U}_{irr}\mathbf{U}_{irr}^T)\mathbf{x}^1(m), \quad m = 1, \dots, M \quad (12)$$

where \mathbf{U}_{irr} is the matrix whose columns are the eigenvectors associated to the non-zero eigenvalues of $\mathbf{B}_1^T\mathbf{B}_1$. If the ratio between the overall energy of the M signal components $\mathbf{x}_{sH}^1(m)$ and that of the signals $\mathbf{x}^1(m)$ is lower than a given threshold, we set $\mathbf{B}_2 = \mathbf{0}$ and stop, otherwise we proceed to estimate \mathbf{B}_2 . To do that, we build a matrix \mathbf{B}_2 as in (3), identifying all polygons in the graph. This matrix can be rewritten as

$$\mathbf{B}_2\mathbf{B}_2^T = \sum_{n=1}^P q_n \mathbf{b}_n \mathbf{b}_n^T \quad (13)$$

where \mathbf{b}_n is the column of \mathbf{B}_2 associated with the n -th 2-cell (polygon), whose entries are all zero except the entries associated with the edges of the n -th cell. The coefficient $q_n \in \{0, 1\}$ is introduced to select the columns of \mathbf{B}_2 for which the total circulation of the observed edge signal along the corresponding 2-cell is minimum. Note that the circulation of the generic component $\mathbf{x}_{sH}^1(m)$ along the boundary of the n -th polygon is $\mathbf{b}_n^T \mathbf{x}_{sH}^1(m)$. If the overall sum of the norm of this circulation is sufficiently small over all set of data, we set $q_n = 1$, i.e. we fill the n -th polygon, otherwise we set $q_n = 0$ and leave the corresponding polygon empty. Mathematically speaking, the inference problem boils down to solving the following problem

$$\begin{aligned} \min_{\mathbf{q} \in \{0,1\}^P} \quad & f_{TV}(\mathbf{q}) := \sum_{n=1}^P \sum_{m=1}^M q_n [\mathbf{b}_n^T \mathbf{x}_{sH}^1(m)]^2 \quad (\mathcal{P}_{MTV}) \\ \text{s.t.} \quad & \|\mathbf{q}\|_0 = q^*, \quad q_n \in \{0, 1\}, \forall n, \end{aligned} \quad (14)$$

where q^* is the number of (filled) 2-cells. Note that the objective function f_{TV} represents the sum of the squared variations (curls) of the edge signals along the 2-cells of the complex.

Although problem \mathcal{P}_{MTV} is non-convex, it can be easily solved in closed form, as in [17]. In fact, defining the non-negative coefficients $d_n = \sum_{m=1}^M [\mathbf{b}_n^T \mathbf{x}_{sH}^1(m)]^2$, the optimal solution of problem \mathcal{P}_{MTV} can be obtained by sorting the coefficients d_n in increasing order and selecting the columns of \mathbf{B}_2 corresponding to the indices of the q^* smallest coefficients d_n . Note that the proposed inference strategy can be extended to planar hollow cell complexes by using the incidence matrix \mathbf{B}_2 in (41) in the supplementary material.

Remark: In problem \mathcal{P}_{MTV} , we need to select the optimal number q^* of filled 2-cells first. However, given the simplicity of the closed form solution, we can compute the solution associated to all values of q^* and then select the q^* value yielding

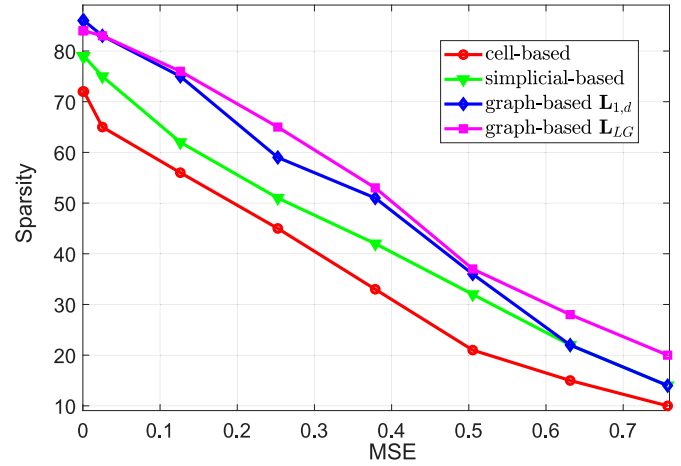


Fig. 5. Sparsity versus mean squared error.

the best sparsity/accuracy trade-off. To check the effectiveness of the proposed approach, we compare the method with the corresponding method used in [17] to infer a simplicial complex from data and with graph-based approaches. The criterion used for the comparison is the trade-off between sparsity of the representation and fitting error. To perform the comparison, given each observed flow vector \mathbf{x}^1 , we find the sparse \mathbf{s}^1 as solution of the following basis pursuit problem [49]:

$$\begin{aligned} \min_{\hat{\mathbf{s}}^1 \in \mathbb{R}^E} \quad & \|\hat{\mathbf{s}}^1\|_1 \quad (\mathcal{P}_B) \\ \text{s.t.} \quad & \|\mathbf{x}^1 - \mathbf{V}\hat{\mathbf{s}}^1\|_F \leq \epsilon \end{aligned} \quad (15)$$

where \mathbf{V} is a dictionary matrix, whose columns are the eigenvectors \mathbf{U}_1 of the Laplacian matrices built using four alternative methods: 1) the cell-based method, where $\mathbf{L}_1 = \mathbf{B}_1^T\mathbf{B}_1 + \mathbf{B}_2\mathbf{B}_2^T$, with \mathbf{B}_2 built using the cell-based method described above; 2) the simplicial-based method, where $\mathbf{L}_1 = \mathbf{B}_1^T\mathbf{B}_1 + \mathbf{B}_2\mathbf{B}_2^T$, with \mathbf{B}_2 built using the simplicial-based method proposed in [17]; 3) the graph-based method, where $\mathbf{L}_{1,d} = \mathbf{B}_1^T\mathbf{B}_1$ incorporates only a purely graph-based representation; 4) a line graph-based method, where the line graph [50] of a graph \mathcal{G} is a graph whose nodes correspond to the edges of \mathcal{G} and its Laplacian matrix is defined as $\mathbf{L}_{LG} = \text{diag}(\mathbf{A}_{LG}\mathbf{1}) - \mathbf{A}_{LG}$ with $\mathbf{A}_{LG} = |\mathbf{L}_{1,d} - 2\mathbf{I}|$. By varying the positive coefficient ϵ in (15), we can explore the trade-off between sparsity and accuracy.

Data Traffic Network. We tested the above four methods on the real traffic data measurements on the German National Research and Education Network operated by the German DFN-Verein (DFN) [51]. The backbone network consists of 50 nodes, 89 links and 39 potential 2-cells. The data traffic is aggregated daily over the month of February 2005 and the data measurements, expressed in Mbit/sec, are observed over the links of the backbone network. In Fig. 5 we draw the sparsity vs. the representation error $\|\mathbf{x}^1 - \mathbf{V}\hat{\mathbf{s}}^1\|_F$ obtained with the above four methods by varying the coefficient ϵ . We can see from Fig. 5 that the cell-based approaches outperforms all other methods in terms of sparsity/accuracy trade-off.

VI. EXTRACTION OF SPACE INVARIANTS (HOMOLOGIES)

If embedded in a two-dimensional (2D) real domain \mathbb{R}^2 , a geometric cell complex typically fills a surface. More generally, in a higher dimension domain \mathbb{R}^m , the complex may fill an m -th dimensional volume. A basic result of algebraic topology is that some of the fundamental invariants of a cell complex are the number of connected components and the number of holes (cavities) appearing at increasing dimensions [43]. When analyzing flows along the edges of a complex, it is of particular interest to identify closed paths, or cycles, and check the flow of signals along these cycles. More specifically, looking for invariants, it is useful to distinguish between two kinds of cycles: the cycles, also named boundaries, that enclose a filled surface (or volume) and the cycles that enclose a hole (or cavity). Detecting cycles surrounding holes is indeed a fundamental step to identify invariants of the structure and identifying invariants is a key step in devising effective geometric deep learning algorithms [52]. Detecting cycles in flow signals finds applications in protein-protein interaction networks [53] to recognize proteins with similar biological functions or to categorize protein into enzyme and not-enzyme structures [54]. The distinction between cycles that are or not boundaries can also be made using simplicial complexes, but in the latter case, for example, we cannot include filled polygons of dimension greater than three and in many applications such a restriction may prevent us from finding cycles with minimal length [55], [56]. The harmonic (edge) signals are (global) circulations that are not representable as a linear combination of curl flows. The analysis of harmonic signals over simplicial complexes has been used to detect global inconsistency in statistical ranking [11], to localize coverage holes in a sensor network represented as a Rips complex [37], and to identify topologically similar trajectories through embedding into the harmonic subspace [19]. The harmonic signals can be represented as a linear combination of a basis of eigenvectors spanning the kernel of \mathbf{L}_1 . Since there is no unique way to identify a basis for such a subspace, we look for the sparsest cycles, i.e. cycles that do not include boundaries of 2-order cells, and that cover a minimum number of edges around holes. More specifically, in this section, we propose a distributed optimization strategy to identify holes over cell complexes from noisy measurements by finding the optimal trade-off between data fitting error and sparsity of the edge signals. We derive the sparsest harmonic component by using an iterative subspace projection optimization method [57] converging to the desired harmonic solution and amenable for a distributed implementation. Then, if our goal is detecting holes we have to remove from the harmonic signal \mathbf{s}_H the curl components that can be expressed as combinations of circulations along the boundaries of the 2-cells. Assume that the noisy edge signal $\mathbf{x}^1 \in \mathbb{R}^E$ is observed over the edges of a cell complex of order 2 with

$$\mathbf{x}^1 = \mathbf{s}_H + \mathbf{n} \quad (16)$$

where \mathbf{n} is an additive noise vector. Our goal is to filter the harmonic component \mathbf{s}_H in order to obtain a sparse representation useful to detect holes by using a distributed algorithm. Then, we have to extract from the observed edge signal \mathbf{x}^1 the harmonic

component while removing the contributions corresponding to curl flows of 2-cells, represented by the signal $\mathbf{s}_{sol} = \mathbf{B}_2 \mathbf{s}_2$. This entails minimizing both the fitting error and the l_1 -norm of the vector $\mathbf{z} = \mathbf{s}_H + \mathbf{B}_2 \mathbf{s}_2$, with the goal of identifying the sparsest edge signal (1-cycle) \mathbf{z} belonging to the kernel of \mathbf{B}_1 . This signal is homologous to \mathbf{s}_H and tightly encircles the holes of the complex. Hence, we find the optimal filtered harmonic signal \mathbf{s}_H and the second-order cell signal \mathbf{s}^2 solving the following optimization problem

$$\begin{aligned} \min_{\mathbf{s}_H, \mathbf{s}^2} \quad & \|\mathbf{x}^1 - \mathbf{s}_H\|^2 + \gamma \|\mathbf{s}_H + \mathbf{B}_2 \mathbf{s}^2\|_1 \quad (\mathcal{Q}) \\ \text{s.t.} \quad & \mathbf{s}_H = \mathbf{W}_h \mathbf{s}_H \end{aligned} \quad (17)$$

where $\mathbf{W}_h = \mathbf{I} - \epsilon \mathbf{L}_1$ and γ is a non-negative coefficient controlling the trade-off between fitting error and sparsity of the recovered edge signal $\mathbf{z} = \mathbf{s}_H + \mathbf{B}_2 \mathbf{s}^2$. The constraint $\mathbf{s}_H = \mathbf{W}_h \mathbf{s}_H$ forces \mathbf{s}_H to belong to the harmonic subspace [57]. To ensure convergence to the harmonic solution of the ensuing iterative solution of the above problem, we choose $0 < \epsilon < 2/\lambda_{max}(\mathbf{L}_1)$ where $\lambda_{max}(\mathbf{L}_1)$ is the maximum eigenvalue of \mathbf{L}_1 . A distributed solution of problem \mathcal{Q} can be obtained proceeding as in [57], by converting the constrained problem \mathcal{Q} into a sequence of unconstrained problems that, at each step k , are formulated as follows:

$$\begin{aligned} \min_{\mathbf{s}_H, \mathbf{s}^2} \quad & f_h(\mathbf{s}_H, \mathbf{s}^2) := \|\mathbf{x}^1 - \mathbf{s}_H\|^2 + \gamma \|\mathbf{s}_H + \mathbf{B}_2 \mathbf{s}^2\|_1 \\ & + \frac{1}{2\mu[k]} \mathbf{s}_H^T (\mathbf{I} - \mathbf{W}_h) \mathbf{s}_H \end{aligned} \quad (18)$$

where $\{\mu[k]\}_k$ is a positive non-increasing sequence forcing the signal \mathbf{s}_H to belong to $\ker(\mathbf{I} - \mathbf{W}_h)$ as $k \rightarrow \infty$. Since we chose $\epsilon > 0$ such that $\mathbf{I} - \mathbf{W}_h = \epsilon \mathbf{L}_1$, from the positive semidefiniteness of \mathbf{L}_1 , it follows that the matrix $\mathbf{I} - \mathbf{W}_h$ is positive semidefinite. Then, problem (18) is a convex non-differentiable problem that can be solved using a (sub)gradient method where the following recursive rules are updated:

$$\begin{aligned} \mathbf{s}^2[k+1] &= \mathbf{s}^2[k] - \mu[k] \partial_{\mathbf{s}^2} f_h(\mathbf{s}_H[k], \mathbf{s}^2[k]) \\ \mathbf{s}_H[k+1] &= \mathbf{s}_H[k] - \mu[k] \partial_{\mathbf{s}_H} f_h(\mathbf{s}_H[k], \mathbf{s}^2[k]) \end{aligned} \quad (19)$$

with

$$\begin{aligned} \partial_{\mathbf{s}^2} f_h(\mathbf{s}_H[k], \mathbf{s}^2[k]) &= \gamma \mathbf{B}_2^T \text{sign}(\mathbf{s}_H[k] + \mathbf{B}_2 \mathbf{s}^2[k]) \\ \partial_{\mathbf{s}_H} f_h(\mathbf{s}_H[k], \mathbf{s}^2[k]) &= -2(\mathbf{x}^1 - \mathbf{s}_H) \\ &+ \gamma \text{sign}(\mathbf{s}_H[k] + \mathbf{B}_2 \mathbf{s}^2[k]). \end{aligned} \quad (20)$$

Note that the updating rules in (19) are amenable for a distributed implementation since the subgradients in (20) can be derived combining 2- or 1-order signals according to the neighborhood relationships captured by the incidence matrix \mathbf{B}_2 . The final distributed algorithm is illustrated in Algorithm 1 where, to ensure convergence, we adopt the nonsummable diminishing step size rule $\mu[k] = a/\sqrt{k}$ with $a > 0$.

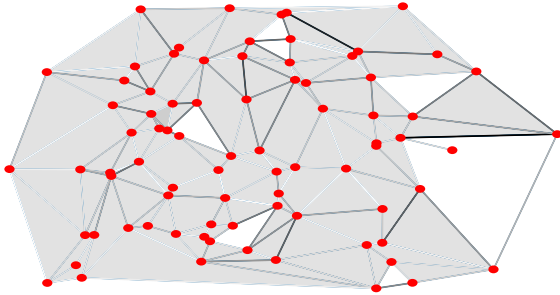
To test the goodness of the proposed harmonic filtering method in identifying minimal-length cycles around holes, we consider the cell complex in Fig. 6(a) composed of $N = 80$ nodes, a number of edges $E = 188$ and 102 polyhedral 2-cells.

Algorithm 1 Distributed Optimization of the Harmonic Filtering

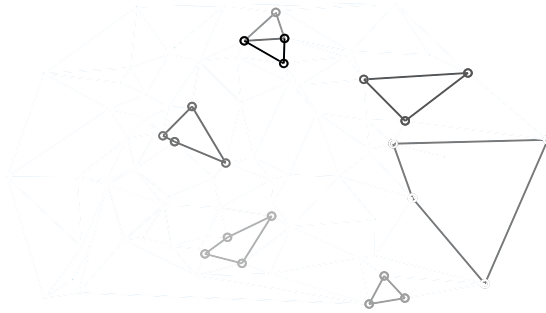
Data: \mathbf{x}^1 . Start with random vectors $\mathbf{s}_H[0] \in \mathbb{R}^{E \times 1}$, $\mathbf{s}^2[0] \in \mathbb{R}^{T \times 1}$, $a = 1.5e^{-3}$, $k = 0$, $\mu[k] = a$.

For each time $k > 0$, repeat

- 1) $\mathbf{s}^2[k+1] = \mathbf{s}^2[k] - \mu[k] \partial_{\mathbf{s}^2} f_h(\mathbf{s}_H[k], \mathbf{s}^2[k])$
- 2) $\mathbf{s}_H[k+1] = \mathbf{s}_H[k] - \mu[k] \partial_{\mathbf{s}_H} f_h(\mathbf{s}_H[k], \mathbf{s}^2[k])$
- 3) $\mu[k+1] = \frac{a}{\sqrt{k+1}}$
- 4) $k = k + 1$



(a) cell complex



(b) recovered edge signal

Fig. 6. Filtering harmonic component.

The covered surface contains 7 holes so that the dimension of the harmonic subspace is equal to 7. The observed edge signal is generated as $\mathbf{x}^1 = \mathbf{U}_H \hat{\mathbf{s}}^1 + \mathbf{n}$ where \mathbf{U}_H are the eigenvectors spanning the kernel of \mathbf{L}_1 , $\hat{\mathbf{s}}^1$ is a random vector with entries uniformly distributed between $[0, 1]$ and the vector \mathbf{n} has zero-mean Gaussian entries with variance $\sigma_n^2 = 0.05$. The signal values are encoded in the gray color associated with the links in Fig. 6(a). The sparse edge signal $\mathbf{s}_H + \mathbf{B}_2 \mathbf{s}^2$ recovered by running Algorithm 1 is reported in Fig. 6(b), by setting the trade-off coefficient $\gamma = 50$. We can observe that the algorithm converges to an edge signal that is non-zero only along the cycles surrounding the holes. This is an example testifying the goodness of the method to identify the (shortest) cycles of the complex.

VII. FILTERING OVER CELL COMPLEXES

The design of filters operating over signals defined over the edges of a graph, exploiting simplicial complex topologies, has been considered recently in [17], [26], [28]. A first introduction to filtering over cell-complexes has been provided in [8]. In [1] we introduced some preliminary results on finite impulse response (FIR) filters over cell complexes. A local filter operating over an edge signal vector \mathbf{s}^1 can be modeled as

$$\mathbf{y}^1 = \mathbf{H} \mathbf{s}^1 \quad (21)$$

where we assume the support of \mathbf{H} given by the (upper and lower) neighbors of each edge. Generalizing the FIR filter design operating over graphs, see e.g. [58], to cell complexes, a polynomial FIR filter based on the first order Laplacian has the following form

$$\mathbf{H} = \sum_{k=0}^L a_k \mathbf{L}_1^k = \alpha_0 \mathbf{I} + \sum_{k=1}^L a_k^I \mathbf{L}_{1,d}^k + \sum_{k=1}^L a_k^s \mathbf{L}_{1,u}^k \quad (22)$$

where $\alpha_0 := a_0^I + a_0^s$ and L is the filter order. In the last equality, we used (5) and the property that, since $\mathbf{B}_1 \mathbf{B}_2 = \mathbf{0}$, we can always write $\mathbf{L}_1^k = \mathbf{L}_{1,d}^k + \mathbf{L}_{1,u}^k$. The above expression shows that filtering edge signals occurs by combining, separately, along lower neighbors (first term) and upper neighbors (second term). In (22), we adopt two sets of coefficients $\{a_k^I\}_{k=1}^L$ and $\{a_k^s\}_{k=1}^L$ to design the filters operating along the lower and upper neighbors. This formulation was already proposed in [26] to design filters operating over simplicial complexes.

A possible way to select the two sets of filter coefficients can be obtained by imposing two spectral masks on the irrotational and solenoidal components. Applying a CFT to the output of the filter, we can write the spectrum of the output as

$$\hat{\mathbf{y}}^1 = \mathbf{U}_1^T \mathbf{y}^1 = \left(\alpha_0 \mathbf{I} + \sum_{k=1}^L a_k^I \mathbf{\Lambda}_{1,d}^k + \sum_{k=1}^L a_k^s \mathbf{\Lambda}_{1,u}^k \right) \hat{\mathbf{s}}^1, \quad (23)$$

where \mathbf{U}_1 are the eigenvectors of \mathbf{L}_1 and $\mathbf{\Lambda}_{1,d}$, $\mathbf{\Lambda}_{1,u}$ are the diagonal matrices with entries the eigenvalues $\lambda_d(i)$, $\lambda_u(i)$ of, respectively, the lower and upper Laplacians. From this expression we can notice that the effect of the filters is to impose on the irrotational and solenoidal frequencies the spectral masks

$$h^I(\mathbf{\Lambda}_{1,d}) = \sum_{k=1}^L a_k^I \mathbf{\Lambda}_{1,d}^k, \quad h^s(\mathbf{\Lambda}_{1,u}) = \sum_{k=1}^L a_k^s \mathbf{\Lambda}_{1,u}^k, \quad (24)$$

whose diagonal entries, given by respectively, $h^I(\lambda_d(i)) = \sum_{k=1}^L a_k^I \lambda_d(i)^k$ and $h^s(\lambda_u(i)) = \sum_{k=1}^L a_k^s \lambda_u(i)^k$ are the frequency responses of the filters. Note that, if the Laplacian \mathbf{L}_1 contains eigenvalues of multiplicity greater than 1, all associated to the same subspace, i.e. to the irrotational, solenoidal or harmonic subspace, the spectral filter is unable to distinguish components belonging to the same subspace. However, if a multiple eigenvalue is associated with both solenoidal and irrotational eigenvectors, the spectral filter can distinguish components within either the solenoidal or irrotational subspace, only if the multiplicity of the eigenvalue within the subspace is equal to one. A relevant example of this situation occurs

when the second order cell complex covers a surface, but it leaves a number of holes $N_h > 0$. In such a case, the multiplicity of the null eigenvalue is exactly N_h [6]. From (23), we can immediately see that the harmonic component, associated with the null eigenvalue, is simply multiplied by a scalar coefficient α_0 . Hence, the filter structure in (23) is not adequate to filter *within* the harmonic subspace. To overcome this issue, we can resort to method proposed in the previous section that operates explicitly on the harmonic component.

Filtering the solenoidal and irrotational components. In this section, we will provide theoretical conditions under which the design of two separate FIR filters based on the lower and upper Laplacians may offer performance gains with respect to a single FIR filter handling jointly the solenoidal and irrotational signals. The solenoidal and irrotational components refer, by construction, to eigenvalues different from zero, then we define the corresponding filters by setting $\alpha_0 = 0$ in (22) and (23).

Given the spectral mask $h(\lambda_{1,i})$ that we wish to implement over the non-zero eigenvalues $\lambda_{1,i}$ of \mathbf{L}_1 , the spectral \mathbf{L}_1 -based filter can be written as

$$\mathbf{H} = \bar{\mathbf{U}}_1 h(\mathbf{\Lambda}_1) \bar{\mathbf{U}}_1^T \quad (25)$$

where $\bar{\mathbf{U}}_1$ contains, in its columns, only the eigenvectors associated to the non-null eigenvalues of \mathbf{L}_1 , whereas $h(\mathbf{\Lambda}_1)$ is a diagonal matrix with entries $h(\lambda_{1,i})$, $i = 1, \dots, n_t$, and $n_t = \text{rank}(\mathbf{L}_1)$. The filter in (25) can be decomposed in the sum of two orthogonal filters associated with the irrotational and solenoidal subspaces, i.e.

$$\mathbf{H} = \mathbf{H}_d + \mathbf{H}_u = \mathbf{U}_d h^I(\mathbf{\Lambda}_d) \mathbf{U}_d^T + \mathbf{U}_u h^s(\mathbf{\Lambda}_u) \mathbf{U}_u^T, \quad (26)$$

where $\mathbf{\Lambda}_d$ and $\mathbf{\Lambda}_u$ are the diagonal matrices whose entries are the non-zero eigenvalues of $\mathbf{L}_{1,d}$ and $\mathbf{L}_{1,u}$, whereas \mathbf{U}_d and \mathbf{U}_u are the matrices containing the associated eigenvectors. Let us now consider the two cases where the solenoidal and irrotational filters are designed using a common set of coefficients or two separate sets.

1) *Common Filtering:* Let us first approximate the filter \mathbf{H} using a FIR filter based on the first order Laplacian given by

$$\hat{\mathbf{H}} = \sum_{k=1}^L a_k \mathbf{L}_1^k. \quad (27)$$

The filter coefficient vector $\mathbf{a} = [a_1, \dots, a_L]^T$ can be found as the solution of a least squares problem. Let us define the matrices

$$\Phi_I = [\lambda_I, \lambda_I^2, \dots, \lambda_I^L], \quad \Phi_s = [\lambda_s, \lambda_s^2, \dots, \lambda_s^L] \quad (28)$$

with λ_I and λ_s the column vectors associated with the diagonal entries of $\mathbf{\Lambda}_d$ and $\mathbf{\Lambda}_u$. Then, using the equalities in (26) and the orthogonality condition $\mathbf{U}_d^T \mathbf{U}_u = \mathbf{0}$, we get

$$\begin{aligned} \|\mathbf{H} - \sum_{k=1}^L a_k \mathbf{L}_1^k\|_F^2 &= \|\mathbf{h}^s - \Phi_s \mathbf{a}\|_F^2 + \|\mathbf{h}^I - \Phi_I \mathbf{a}\|_F^2 \\ &= \|\mathbf{h} - \Phi_t \mathbf{a}\|_F^2 \end{aligned} \quad (29)$$

where $\mathbf{h}^s = \text{diag}(h^s(\mathbf{\Lambda}_u))$, $\mathbf{h}^I = \text{diag}(h^I(\mathbf{\Lambda}_d))$, $\mathbf{h} = [\mathbf{h}^s; \mathbf{h}^I]$ and $\Phi_t = [\Phi_s; \Phi_I]$. Hence, in the spectral domain the least squares problem we need to solve is

$$\min_{\mathbf{a} \in \mathbb{R}^L} f(\mathbf{a}) := \|\mathbf{h} - \Phi_t \mathbf{a}\|_F^2 \quad (30)$$

whose closed form solution is given by $\mathbf{a} = \Phi_t^\dagger \mathbf{h}$ where Φ_t^\dagger is the Moore-Penrose pseudo-inverse of Φ_t .

2) *Independent Filter Design:* The FIR filter design illustrated above is the result of a parsimonious choice of the number of filter coefficients, but may be often unable to well approximate different masks for the solenoidal and irrotational components. To improve the situation, recalling (22) and (23), it is worth to design the two FIR filters separately, at the slightly increased cost of using two sets of coefficients. More specifically, given the linear filters \mathbf{H}_u and \mathbf{H}_d , our goal now is to find the optimal FIR filters $\hat{\mathbf{H}}_u$, $\hat{\mathbf{H}}_d$ approximating them such that

$$\hat{\mathbf{H}}_u = \sum_{k=1}^L a_k^s \mathbf{L}_{1,u}^k, \quad \hat{\mathbf{H}}_d = \sum_{k=1}^L a_k^I \mathbf{L}_{1,d}^k. \quad (31)$$

Proceeding exactly as before, the two filters can be found by solving the following least squares problems in the spectral domain:

$$\min_{\mathbf{a}^s \in \mathbb{R}^L} f_s(\mathbf{a}^s) := \|\mathbf{h}^s - \Phi_s \mathbf{a}^s\|_F^2, \quad (32)$$

$$\min_{\mathbf{a}^I \in \mathbb{R}^L} f_I(\mathbf{a}^I) := \|\mathbf{h}^I - \Phi_I \mathbf{a}^I\|_F^2. \quad (33)$$

The optimal coefficients solving problems (32) and (33) can be derived in closed form.

Given the two optimization problems in (32) and (33), let us denote with L , n_d and n_u , respectively, the filter length and the number of distinct non zero eigenvalues of $\mathbf{L}_{1,d}^1$ and $\mathbf{L}_{1,u}^1$. Then, as proved in Appendix B of the supplementary material, it holds:

a) the optimal filter coefficients solving (32), (33) are

$$\mathbf{a}^{s*} = \Phi_s^\dagger \mathbf{h}^s, \quad \mathbf{a}^{I*} = \Phi_I^\dagger \mathbf{h}^I. \quad (34)$$

b) if $L \geq n_d$ and $L \geq n_u$, then the pseudoinverse solutions in (34) are such that $f_s(\mathbf{a}^{s*}) = f_I(\mathbf{a}^{I*}) = 0$. Furthermore, if $L = n_u$ and/or $L = n_d$ then the solutions of (32) and/or (33) are unique.

Even though the independent filters design in general leads performance gains with respect to a single filter due to the higher degree of freedoms, we indeed observed that there are some particular cases where this improvement does not occur. To better understand when this happens, in the following theorem we state the theoretical conditions under which the design of independent FIR filters for the solenoidal and irrotational signals attains better performance with respect to a single \mathbf{L}_1 -based filter. Let us define

$$\Phi_d = [\text{vec}(\mathbf{L}_{1,d}^1), \dots, \text{vec}(\mathbf{L}_{1,d}^L)], \quad \Phi_u = [\text{vec}(\mathbf{L}_{1,u}^1), \dots, \text{vec}(\mathbf{L}_{1,u}^L)]. \quad (35)$$

Denote with $\mathcal{R}(\mathbf{A})$ the span of the columns of the matrix \mathbf{A} , and with $\mathbf{P}_{\mathcal{R}(\mathbf{A})} = \mathbf{A} \mathbf{A}^\dagger$, $\mathbf{P}_{\mathcal{R}(\mathbf{A})^\perp} = \mathbf{I} - \mathbf{A} \mathbf{A}^\dagger$ the projections, respectively, onto $\mathcal{R}(\mathbf{A})$ and onto its complementary subspace.

Proposition 1: Let $f(\mathbf{a}^*)$, $f_s(\mathbf{a}^{s*})$ and $f_I(\mathbf{a}^{I*})$ be the optimal solutions of the convex optimization problems in (30),

(32) and (33), respectively. Define the vector $\mathbf{y} = (\mathbf{P}_{\mathcal{R}(\Phi_d)} + \mathbf{P}_{\mathcal{R}(\Phi_u)})\text{vec}(\mathbf{H})$. Then, it holds:

a) if the system $\Phi_t \mathbf{a} = \mathbf{h}$ is inconsistent and the matrices Φ_d and Φ_u are full column rank, then:

i) If $\mathbf{P}_{\mathcal{R}(\Phi_d + \Phi_u)} \mathbf{y} = \mathbf{0}$ the disjoint and joint filtering strategies are equivalent, since we have:

$$f_I(\mathbf{a}^{I^*}) + f_s(\mathbf{a}^{s^*}) = f(\mathbf{a}^*); \quad (36)$$

ii) If $\mathbf{P}_{\mathcal{R}(\Phi_d + \Phi_u)} \mathbf{y} \neq \mathbf{0}$, the disjoint filtering strategy leads smaller filter estimation errors since it holds:

$$f_I(\mathbf{a}^{I^*}) + f_s(\mathbf{a}^{s^*}) < f(\mathbf{a}^*); \quad (37)$$

b) if the system $\Phi_t \mathbf{a} = \mathbf{h}$ is consistent, then the disjoint and joint filtering strategies are equivalent since we get

$$f_I(\mathbf{a}^{I^*}) = f_s(\mathbf{a}^{s^*}) = f(\mathbf{a}^*) = 0. \quad (38)$$

Proof: See Appendix C in the supplementary material. ■

From Prop. 1, we can state that, in terms of estimation error, the advantage achievable using the separate design depends on the structure of the desired filter \mathbf{H} . More specifically, by keeping the length L of the approximating FIR filters fixed, if the system $\Phi_t \mathbf{a} = \mathbf{h}$ is consistent, then the common or separate design would provide the same accuracy. If, instead, the system is inconsistent and the matrices Φ_d and Φ_u are full column rank, then we compute first the projection \mathbf{y} of $\text{vec}(\mathbf{H})$ onto the subspace $\mathcal{R}(\Phi_d) + \mathcal{R}(\Phi_u)$. If \mathbf{y} has a component falling outside the subspace $\mathcal{R}(\Phi_d + \Phi_u)$, then the separate design of the solenoidal and irrotational parts yields a lower estimation error.

As a first numerical test, we compare the performance of our filtering strategy with the approach proposed in [26], adapted to filtering signals defined over cell complexes by simply replacing \mathbf{B}_2 with the matrix in (3) capturing the upper incidence relations of a cell complex. Our optimal FIR filters are derived by solving the least squares problem in (32) and (33), where we set $\alpha_0 = 0$. Indeed, albeit the difference between our approach and that of [26] is minimum, because we simply set $\alpha_0 = 0$ a priori, while in [26] the choice of α_0 is part of the optimization, we will show that there is a clear advantage in terms of accuracy. As a numerical example, we consider the filtering of the solenoidal component over a cell complex composed of $N = 80$ vertices, $E = 188$ edges and 96 cells of order 2 with $\dim(\ker(\mathbf{L}_1)) = 13$. The input signal \mathbf{s}^1 is the sum of a solenoidal plus a harmonic signal, generated as $\mathbf{s}_{sol}^1 = \mathbf{U}_u \mathbf{x}_{sol}$ and $\mathbf{s}_H^1 = \mathbf{U}_H \mathbf{x}_H$, respectively, where \mathbf{U}_H is a set of orthonormal vectors spanning the harmonic subspace and the vectors $\mathbf{x}_{sol}, \mathbf{x}_H$ have random i.i.d. entries with uniform p.d.f. between $[0, a]$. To check the impact of the harmonic component, we keep the variance of the solenoidal part σ_{sol}^2 fixed, by setting $a = 10$, and we increase the variance σ_H^2 of the harmonic component. The frequency response of the solenoidal filter is chosen equal to

$$h^s(\boldsymbol{\lambda}_s; \alpha_s, b_s) = \frac{\beta_s}{1 + \exp(\alpha_s(\boldsymbol{\lambda}_s - b_s))} \quad (39)$$

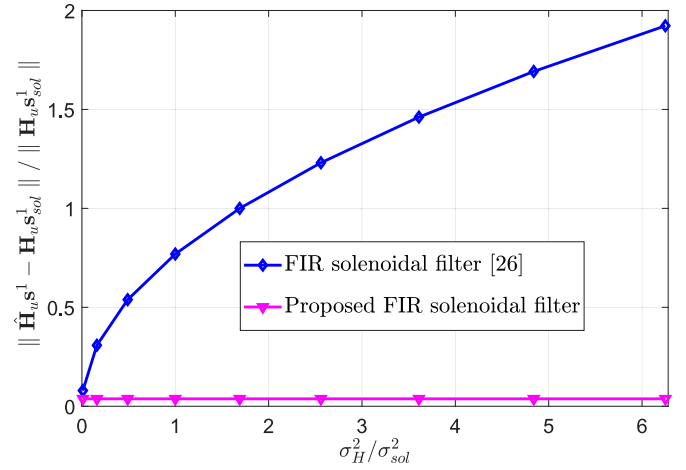


Fig. 7. NMSE of the filtered solenoidal signal vs the ratio $\frac{\sigma_H^2}{\sigma_{sol}^2}$.

with $\beta_s = 0.8$, $\alpha_s = b_s = 3$, and the length of the FIR filters is $L = 11$. In Fig. 7 we plot the average ratio $\frac{\|\hat{\mathbf{H}}_u \mathbf{s}^1 - \mathbf{H}_u \mathbf{s}_{sol}^1\|}{\|\mathbf{H}_u \mathbf{s}_{sol}^1\|}$, i.e. the normalized mean squared error between the signal at the output of the solenoidal FIR filter taking as input \mathbf{s}^1 and the output of the ideal filter \mathbf{H}_u , built as in (26) using the spectral mask in (39), with input signal \mathbf{s}_{sol}^1 , versus the ratio $\sigma_H^2 / \sigma_{sol}^2$. The numerical results are averaged over 50 random signal realizations. We can notice from Fig. 7 that, while our filter is insensitive to the harmonic component, the filtering strategy proposed in [26] is affected by the harmonic component.

To investigate the benefits of using cell-based against simplicial-based filters over real data, we considered the transportation network and the associated real traffic from the city of Anaheim [59], in the year 1992. This network is composed of $N = 416$ nodes, $E = 914$ links and 208 polygons. The flow f_{ij} over each link (i, j) is bidirectional so we define the flow signal over edge (i, j) as $s_{ij} = (f_{ij} + f_{ji})/2$. Then, we built the Laplacian matrices of the simplicial complex and of the cell complexes associated to the graph and we computed the associated eigenvector matrices \mathbf{U}_{sc} and \mathbf{U}_1 , respectively. The two complexes have been built by filling all possible $T = 54$ triangles in the SC case and by inferring the cell complex with 198 polygons in the CC case. Then, we projected the observed edge flow \mathbf{s}^1 onto the two sets of eigenvectors \mathbf{U}_{sc} and \mathbf{U}_1 , giving rise to the two vectors $\hat{\mathbf{s}}_{sc}^1 = \mathbf{U}_{sc}^T \mathbf{s}^1$ and $\hat{\mathbf{s}}_{cc}^1 = \mathbf{U}_1^T \mathbf{s}^1$. We then build two diagonal selection matrices \mathbf{D}_{sc} and \mathbf{D}_1 with entries equal to 1 in the positions associated with the largest B frequency coefficients, respectively of $\hat{\mathbf{s}}_{sc}^1$ and $\hat{\mathbf{s}}_{cc}^1$. We added a random zero-mean Gaussian noise with unit variance to the observed flow and then we filtered the resulting signal using the two filters $\mathbf{H}_{sc} = \mathbf{U}_{sc} \mathbf{D}_{sc} \mathbf{U}_{sc}^T$ and $\mathbf{H}_1 = \mathbf{U}_1 \mathbf{D}_1 \mathbf{U}_1^T$. In Fig. 8 we report the SNR improvement factor $\text{SNR}_{out} / \text{SNR}_{in}$, i.e. the ratio between the output and the input Signal-to-Noise ratio, versus the signal bandwidth B . We can observe that the use of CC-based filter yields a substantial performance gain compared to the SC-based filter, due to its better capability of filling a surface using polygons of arbitrary dimension.

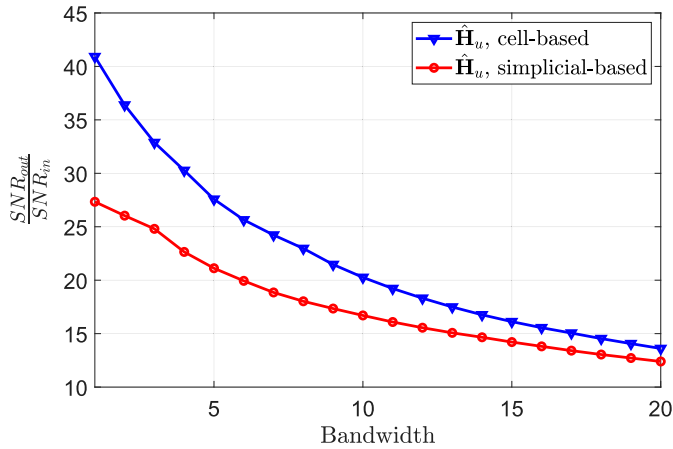


Fig. 8. SNR improvement factor vs the signal bandwidth.

VIII. CONCLUSION

In this paper, we proposed an algebraic framework for analyzing signals represented over cell complexes by showing the significant advantages of using cell spaces compared to simplicial complexes. We introduced planar CC whose cells can contain holes, as a suitable framework to handle images or (approximately) piecewise signals. We provided efficient methods to infer a cell complex from flow signals and investigated alternative strategies to filter their harmonic, solenoidal, and irrotational components. We derived the theoretical conditions guaranteeing a performance gain associated with the independent design of solenoidal and irrotational FIR filters, as opposed to a common FIR filter design. Further investigations should address the incorporation of the proposed filtering and inference capabilities within topological deep learning frameworks, such as graph neural networks, and a more in-depth analysis of the image coding capabilities of the new (hollow) cell complexes.

REFERENCES

- [1] S. Sardellitti, S. Barbarossa, and L. Testa, "Topological signal processing over cell complexes," in *Proc. IEEE Asilomar Conf. Signals, Syst. Comput.*, Nov. 2021, pp. 1558–1562.
- [2] G. Carlsson, "Topology and data," *Bull. Amer. Math. Soc.*, vol. 46, no. 2, pp. 255–308, 2009.
- [3] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 83–98, May 2013.
- [4] R. Lambiotte, M. Rosvall, and I. Scholtes, "From networks to optimal higher-order models of complex systems," *Nat. Phys.*, vol. 15, no. 4, pp. 313–320, 2019.
- [5] C. Berge, *Hypergraphs: Combinatorics of Finite Sets*. Amsterdam, The Netherlands: Elsevier, 1989.
- [6] J. R. Munkres, *Elements of Algebraic Topology*. Boca Raton, FL, USA: CRC Press, 2018.
- [7] A. Hatcher, *Algebraic Topology*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
- [8] L. J. Grady and J. R. Polimeni, *Discrete Calculus: Applied Analysis on Graphs for Computational Science*. New York, NY, USA: Springer-Verlag, 2010.
- [9] M. Hajij et al., "Topological deep learning: Going beyond graph data," 2022, [arXiv:2206.00606v3](https://arxiv.org/abs/2206.00606v3).
- [10] H. Edelsbrunner and J. Harer, "Persistent homology—A survey," *Contemp. Math.*, vol. 453, no. 6, pp. 257–282, 2008.
- [11] X. Jiang, L. H. Lim, Y. Yao, and Y. Ye, "Statistical ranking and combinatorial Hodge theory," *Math. Program.*, vol. 127, no. 1, pp. 203–244, 2011.
- [12] A. Muhammad and M. Egerstedt, "Control using higher order Laplacians in network topologies," in *Proc. 17th Int. Symp. Math. Theory Netw. Syst.*, 2006, pp. 1024–1038.
- [13] T. Roman, A. Nayyeri, B. T. Fasy, and R. Schwartz, "A simplicial complex-based approach to unmixing tumor progression data," *BMC Bioinf.*, vol. 16, no. 1, 2015, Art. no. 254.
- [14] M. Andjelković, B. Tadić, and R. Melnik, "The topology of higher-order complexes associated with brain hubs in human connectomes," *Sci. Rep.*, vol. 10, no. 1, pp. 1–10, 2020.
- [15] M. Robinson, *Topological Signal Processing*. New York, NY, USA: Springer-Verlag, 2016.
- [16] H. Krim and A. B. Hamza, *Geometric Methods in Signal and Image Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2015.
- [17] S. Barbarossa and S. Sardellitti, "Topological signal processing over simplicial complexes," *IEEE Trans. Signal Process.*, vol. 68, pp. 2992–3007, 2020.
- [18] S. Barbarossa and S. Sardellitti, "Topological signal processing: Making sense of data building on multiway relations," *IEEE Signal Process. Mag.*, vol. 37, no. 6, pp. 174–183, Nov. 2020.
- [19] M. T. Schaub, Y. Zhu, J.-B. Seby, T. M. Roddenberry, and S. Segarra, "Signal processing on higher-order networks: Livin' on the edge... and beyond," *Signal Process.*, vol. 187, 2021, Art. no. 108149.
- [20] D. Lee and K.-H. Cho, "Topological estimation of signal flow in complex signaling networks," *Sci. Rep.*, vol. 8, no. 1, pp. 1–11, 2018.
- [21] J. Billings, M. Saggat, J. Hlinka, S. Keilholz, and G. Petri, "Simplicial and topological descriptions of human brain dynamics," *Netw. Neurosci.*, vol. 5, no. 2, pp. 549–568, Jun. 2021.
- [22] K. K. Leung, W. A. Massey, and W. Whitt, "Traffic models for wireless communication networks," *IEEE J. Sel. Areas Commun.*, vol. 12, no. 8, pp. 1353–1364, Oct. 1994.
- [23] M. Rosvall and C. T. Bergstrom, "Maps of random walks on complex networks reveal community structure," *Proc. Natl. Acad. Sci.*, vol. 105, no. 4, pp. 1118–1123, 2008.
- [24] M. T. Schaub, A. Benson, P. Horn, G. Lippner, and A. Jadbabaie, "Random walks on simplicial complexes and the normalized Hodge 1-Laplacian," *SIAM Rev.*, vol. 62, no. 2, pp. 353–391, 2020.
- [25] J. Jia, M. T. Schaub, S. Segarra, and A. R. Benson, "Graph-based semi-supervised & active learning for edge flows," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2019, pp. 761–771.
- [26] M. Yang, E. Isufi, M. T. Schaub, and G. Leus, "Finite impulse response filters for simplicial complexes," in *Proc. 29th Eur. Signal Process. Conf. (EUSIPCO)*, 2021, pp. 2005–2009.
- [27] C. Bodnar et al., "Weisfeiler and Lehman go topological: Message passing simplicial networks," in *Proc. Int. Conf. Mach. Learn. (PMLR)*, 2021, pp. 1026–1037.
- [28] S. Ebli, M. Defferrard, and G. Spreemann, "Simplicial neural networks," in *presented at the NeurIPS 2020 Workshop TDA and Beyond*, 2020.
- [29] T. M. Roddenberry, N. Glaze, and S. Segarra, "Principled simplicial neural networks for trajectory prediction," in *Proc. Int. Conf. Mach. Learn.*, PMLR, 2021, pp. 9020–9029.
- [30] C. Battiloro, L. Testa, L. Giusti, S. Sardellitti, P. Di Lorenzo, and S. Barbarossa, "Generalized simplicial attention neural networks," 2023, [arXiv:2309.02138](https://arxiv.org/abs/2309.02138).
- [31] T. M. Roddenberry, M. T. Schaub, and M. Hajij, "Signal processing on cell complexes," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2022, pp. 8852–8856.
- [32] M. Hajij et al., "Higher-order attention networks," 2022, [arXiv:2206.00606](https://arxiv.org/abs/2206.00606).
- [33] L. Giusti, C. Battiloro, L. Testa, P. Di Lorenzo, S. Sardellitti, and S. Barbarossa, "Cell attention networks," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Australia, Jun. 2023, pp. 1–8.
- [34] R. Balestrieri, R. Cosentino, B. Aazhang, and R. Baraniuk, "The geometry of deep networks: Power diagram subdivision," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 15832–15841.
- [35] A. Paoluzzi et al., "Topological computing of arrangements with (co)chains," *ACM Trans. Spatial Algorithms Syst. (TSAS)*, vol. 7, no. 1, pp. 1–29, 2020.
- [36] W. Huhnt, M. Sternal, and P. J. Pahl, "Modeling bounded and unbounded space with polyhedra: Topology and operators for manifold cell complexes," *Adv. Eng. Inf.*, vol. 54, 2022, Art. no. 101790.
- [37] A. Tahbaz-Salehi and A. Jadbabaie, "Distributed coverage verification in sensor networks without location information," *IEEE Trans. Autom. Control*, vol. 55, no. 8, pp. 1837–1849, Aug. 2010.

- [38] R. Klette, "Cell complexes through time," in *Proc. Vis. Geometry IX*, vol. 4117, 2000, pp. 134–145.
- [39] B. Schröder, *Ordered Sets: An Introduction With Connections From Combinatorics to Topology*, 2nd ed. New York, NY, USA: Springer-Verlag, 2003.
- [40] P. S. Alexandroff, "Diskrete Räume," *Math. Sbornik (N.S.)*, vol. 2 (44), no. 3, pp. 501–518, 1937.
- [41] J. A. Barmak, *Algebraic Topology of Finite Topological Spaces and Applications*, vol. 2032. New York, NY, USA: Springer-Verlag, 2011.
- [42] J. Hansen and R. Ghrist, "Toward a spectral theory of cellular sheaves," *J. Appl. Comput. Topol.*, vol. 3, no. 4, pp. 315–358, 2019.
- [43] J. R. Munkres, *Topology*. Englewood Cliffs, NJ, USA: Prentice-Hall, Sep. 2000.
- [44] D. Horak and J. Jost, "Spectra of combinatorial Laplace operators on simplicial complexes," *Adv. Math.*, vol. 244, pp. 303–336, 2013.
- [45] L.-H. Lim, "Hodge Laplacians on graphs," in *Geometry and Topology in Statistical Inference, Proc. Symp. Appl. Math.*, vol. 76, S. Mukherjee, Ed., Providence, RI, USA: AMS, 2015.
- [46] J. Gallier, *Discrete Mathematics*. New York, NY, USA: Springer-Verlag, 2011.
- [47] G. Mateos, S. Segarra, A. G. Marques, and A. Ribeiro, "Connecting the dots: Identifying network structure via graph signal processing," *IEEE Signal Process. Mag.*, vol. 36, no. 3, pp. 16–43, May 2019.
- [48] S. Sardellitti, S. Barbarossa, and P. D. Lorenzo, "Graph topology inference based on sparsifying transform learning," *IEEE Trans. Signal Process.*, vol. 67, no. 7, pp. 1712–1727, Apr. 2019.
- [49] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.
- [50] C. Godsil and G. F. Royle, *Algebraic Graph Theory*, vol. 207. New York, NY, USA: Springer-Verlag, 2001.
- [51] S. Orlowski, M. Pióro, A. Tomaszewski, and R. Wessälly, "SNDlib 1.0—Survivable Network Design Library," *Networks*, vol. 55, no. 3, pp. 276–286, 2010.
- [52] M. M. Bronstein, J. Bruna, T. Cohen, and P. Velicković, "Geometric deep learning: Grids, groups, graphs, geodesics, and gauges," 2021, *arXiv:2104.13478*.
- [53] D. Davis et al., "Topology-function conservation in protein–protein interaction networks," *Bioinformatics*, vol. 31, no. 10, pp. 1632–1639, 2015.
- [54] K. M. Borgwardt et al., "Protein function prediction via graph kernels," *Bioinformatics*, vol. 21, no. suppl_1, pp. i47–i56, 2005.
- [55] C. Bodnar et al., "Weisfeiler and Lehman go cellular: CW networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 2625–2640.
- [56] D. Horak, S. Maletić, and M. Rajković, "Persistent homology of complex networks," *J. Stat. Mech.: Theory Exp.*, vol. 2009, no. 3, 2009, Art. no. P03034.
- [57] P. Di Lorenzo, S. Barbarossa, and S. Sardellitti, "Distributed signal processing and optimization based on in-network subspace projections," *IEEE Trans. Signal Process.*, vol. 68, pp. 2061–2076, 2020.
- [58] S. Segarra, A. G. Marques, and A. Ribeiro, "Optimal graph-filter design and applications to distributed linear network operators," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 4117–4131, Aug. 2017.
- [59] B. Stabler, H. Bar-Gera, and E. Sall, "Transportation networks for research," GitHub, Inc., San Francisco, CA, USA, 2016. [Online]. Accessed: Nov. 15, 2021. <https://github.com/bstabler/TransportationNetworks>



Stefania Sardellitti (Senior Member, IEEE) received the M.Sc. degree in electronic engineering from Sapienza University of Rome, Italy, in 1998, and the Ph.D. degree in electrical and information engineering from the University of Cassino, Italy, in 2005. From 2005 to 2019, she was an Appointed Professor of digital communications with the University of Cassino, Italy. Currently, she is a Research Fellow with the Department of Information Engineering, Electronics and Telecommunications, University of Rome, Sapienza, Italy. She received the 2014 and the 2020 IEEE Best Paper Awards from the IEEE Signal Processing Society. Her current research interests are in the area of topological signal processing, mobile edge computing, and distributed optimization.



Sergio Barbarossa (Life Fellow, IEEE) is a Full Professor with Sapienza University of Rome and a Senior Research Fellow with Sapienza School for Advanced Studies (SSAS). He is a EURASIP Fellow. He received the Technical Achievements Award from the European Association for Signal Processing (EURASIP) society, in 2010, and the IEEE Best Paper Awards from the IEEE Signal Processing Society, in the years 2000, 2014, and 2020. He served as an IEEE Distinguished Lecturer from 2013 to 2014. He has been the scientific coordinator of several European projects. He is currently the Coordinator of a national project named Netwin (Network Intelligence). His main current research interests include topological signal processing and learning, generative models for semantic and goal-oriented communications, 6G networks, and distributed edge machine learning.

Open Access funding provided by 'Università degli Studi di Roma "La Sapienza" 2' within the CRUI CARE Agreement