

Topological Signal Processing over Simplicial Complexes

Sergio Barbarossa, *Fellow, IEEE*, Stefania Sardellitti, *Member, IEEE*

Abstract—The goal of this paper is to establish the fundamental tools to analyze signals defined over a topological space, i.e. a set of points along with a set of neighborhood relations. This setup does not require the definition of a metric and then it is especially useful to deal with signals defined over non-metric spaces. We focus on signals defined over simplicial complexes. Graph Signal Processing (GSP) represents a special case of Topological Signal Processing (TSP), referring to the situation where the signals are associated only with the vertices of a graph. Even though the theory can be applied to signals of any order, we focus on signals defined over the edges of a graph and show how building a simplicial complex of order two, i.e. including triangles, yields benefits in the analysis of edge signals. After reviewing the basic principles of algebraic topology, we derive a sampling theory for signals of any order and emphasize the interplay between signals of different order. Then we propose a method to infer the topology of a simplicial complex from data. We conclude with applications to traffic analysis over wireless networks and to the processing of discrete vector fields to illustrate the benefits of the proposed methodologies.

Index Terms—Algebraic topology, graph signal processing, topology inference.

I. INTRODUCTION

Historically, signal processing has been developed for signals defined over a metric space, typically time or space. More recently, there has been a surge of interest to deal with signals that are not necessarily defined over a metric space. Examples of particular interest are biological networks, social networks, etc. The field of graph signal processing (GSP) has recently emerged as a framework to analyze signals defined over the vertices of a graph [2], [3]. A graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ is a simple example of topological space, composed of a set of elements (vertices) \mathcal{V} and a set of edges \mathcal{E} representing *pairwise* relations. However, notwithstanding their enormous success, graph-based representations are not always able to capture all the information present in interconnected systems in which the complex interactions among the system constitutive elements cannot be reduced to pairwise interactions, but require a *multiway* description, as suggested in [4]–[9].

To establish a general framework to deal with complex interacting systems, it is useful to start from a topological space, i.e. a set of elements \mathcal{V} , along with an ensemble of *multiway* relations, represented by a set \mathcal{S} containing subsets

of various cardinality, of the elements of \mathcal{V} . The structure $\mathcal{H}(\mathcal{V}, \mathcal{S})$ is known as a *hypergraph*. In particular, a class of hypergraphs that is particularly appealing for its rich algebraic structure is given by *simplicial complexes*, whose defining feature is the inclusion property stating that if a set \mathcal{A} belongs to \mathcal{S} , then all subsets of \mathcal{A} also belong to \mathcal{S} [10]. Restricting the attention to simplicial complexes is a limitation with respect to a hypergraph model. Nevertheless, simplicial complex models include many cases of interest and, most important, their rich algebraic structure makes possible to derive tools that are very useful for analyzing signals over the complex. Learning simplicial complexes representing the complex interactions among sets of elements has been already proposed in brain network analysis [6], neuronal morphologies [11], co-authorship networks [12], collaboration networks [13], [14], tumor progression analysis [15]. More generally, the use of algebraic topology tools for the extraction of information from data is not new: The framework known as *Topological Data Analysis* (TDA), see e.g. [16], has exactly this goal. Interesting applications of algebraic topology tools have been proposed to control systems [17], statistical ranking from incomplete data [18], [19], distributed coverage control of sensor networks [20]–[22], wheeze detection [23]. One of the fundamental tools of TDA is the analysis of persistent homologies extracted from data [24], [25]. Topological methods to analyze signals and images are also the subjects of the two books [26] and [27].

The goal of our paper is to establish a fundamental framework to analyze signals defined over a simplicial complex. Our approach is complementary to TDA: Rather than focusing, like TDA, on the properties of the simplicial complex extracted from data, *we focus on the properties of signals defined over a simplicial complex*. Our approach includes GSP as a particular case: While GSP focuses on the analysis of signals defined over the vertices of a graph, topological signal processing (TSP) considers signals defined over simplices of various order, i.e. signals defined over nodes, edges, triangles, etc. Relevant examples of edge signals are flow signals, like blood flow between different areas of the brain [28], data traffic over communication links [29], regulatory signals in gene regulatory networks [30], where it was shown that a dysregulation of these regulatory signals is one of the causes of cancer [30]. Examples of signals defined over triplets are co-authorship networks, where a (filled) triangle indicates the presence of papers co-authored by the authors associated to its three vertices [12] and the associated signal value counts the number of such publications. Further examples of even higher order structures are analyzed in [8], with the goal of predicting higher-order links. There are previous works dealing with the

The authors are with the Department of Information Engineering, Electronics, and Telecommunications, Sapienza University of Rome, Via Eudossiana 18, 00184, Rome, Italy. E-mails: {sergio.barbarossa; stefania.sardellitti}@uniroma1.it. This work has been supported by H2020 EU/Taiwan Project 5G CONNI, Nr. AMD-861459-3 and by MIUR, under the PRIN Liquid-Edge contract. Some preliminary results of this work were presented at the 2018 IEEE Workshop in Data Science [1].

analysis of edge signals, like [31]–[34]. More specifically, in [31] the authors introduced a class of filters to analyze edge signals based on the edge-Laplacian matrix [35]. A semi-supervised learning method for learning edge flows was also suggested in [32], where the authors proposed filters highlighting both divergence-free and curl-free behaviors. Other works analyzed edge signals using a line-graph transformation [33], [34]. Random walks evolving over simplicial complexes have been analyzed in [36], [37], [38]. In [36], random walks and diffusion process over simplicial complexes are introduced, while in [38] the authors focused on the study of diffusion processes on the edge-space by generalizing the well-known relationship between the normalized graph Laplacian operator and random walks on graphs.

Building a representation based on a simplicial complex is a straightforward generalization of a graph-based representation. Given a set of time series, it is well known that graph-based representations are very useful to capture relevant correlations or causality relations present between different signals [39], [40]. In such a case, each time series is associated to a node of a graph, and the presence of an edge is an indicator of the relations between the signals associated to its endpoints. As a direct generalization, if we have signals defined over the edges of a graph, capturing their relations requires inferring the presence of triangles associating triplets of edges.

In summary, our main contributions are listed below:

- 1) *we show how to derive a real-valued function capturing triple-wise relations among data, to be used as a regularization function in the analysis of edge signals;*
- 2) *we derive a sampling theory defining the conditions for the recovery of high order signals from a subset of observations, highlighting the interplay between signals of different order;*
- 3) *we propose inference algorithms to extract the structure of the simplicial complex from high order signals;*
- 4) *we apply our edge signal processing tools to the analysis of vector fields, with a specific attention to the recovery of the RiboNucleic Acid (RNA) velocity vector field, useful to predict the evolution of a cell behavior [41].*

We presented some preliminary results of our work in [1]. Here, we extend the work of [1], deriving a sampling theory for signals defined over complexes of various order, proposing new inference methods, more robust against noise, and showing applications to the analysis of wireless data traffic and to the estimation of the RNA velocity vector field.

The paper is organized as follows. Section II recalls the main algebraic principles that will form the basis for the derivation of the signal processing tools carried out in the ensuing sections. In Section III, we will first recall the eigenvectors properties of higher-order Laplacian and the Hodge decomposition. Then, we derive the real-valued extension of an edge set function, capturing triple-wise relations among the elements of a discrete set, which will be used to design unitary bases to represent edge signals. In Section IV, we illustrate some methods to recover the edge signal components from noisy observations. Section V provides theoretical conditions to recover an edge signal from a subset of samples, highlighting the interplay between signals defined over structures

of different order. In Section VI, we illustrate how to use edge signal processing to filter discrete vector fields. Then, in Section VII, we propose some methods to infer the simplicial complex structure from noisy observations, illustrating their performance over both synthetic and real data. Finally, in Section VIII we draw some conclusions.

II. REVIEW OF ALGEBRAIC TOPOLOGY TOOLS

In this section we recall the basic principles of algebraic topology [10] and discrete calculus [42], as they will form the background required for deriving the basic signal processing tools to be used in later sections. We follow an algebraic approach that is accessible also to readers with no specific background on algebraic topology.

A. Discrete domains: Simplicial complexes

Given a finite set $\mathcal{V} \triangleq \{v_0, \dots, v_{N-1}\}$ of N points (vertices), a k -simplex σ_i^k is an unordered set $\{v_{i_0}, \dots, v_{i_k}\}$ of $k+1$ points with $0 \leq i_j \leq N-1$, for $j = 0, \dots, k$, and $v_{i_j} \neq v_{i_n}$ for all $i_j \neq i_n$. A *face* of the k -simplex $\{v_{i_0}, \dots, v_{i_k}\}$ is a $(k-1)$ -simplex of the form $\{v_{i_0}, \dots, v_{i_{j-1}}, v_{i_{j+1}}, \dots, v_{i_k}\}$, for some $0 \leq j \leq k$. Every k -simplex has exactly $k+1$ faces. An *abstract simplicial complex* \mathcal{X} is a finite collection of simplices that is closed under inclusion of faces, i.e., if $\sigma_i \in \mathcal{X}$, then all faces of σ_i also belong to \mathcal{X} . The order (or dimension) of a simplex is one less than its cardinality. Then, a vertex is a 0-dimensional simplex, an edge has dimension 1, and so on. The dimension of a simplicial complex is the largest dimension of any of its simplices. A graph is a particular case of an abstract simplicial complex of order 1, containing only simplices of order 0 (vertices) and 1 (edges).

If the set of points is embedded in a real space \mathbb{R}^D of dimension D , we can associate a *geometric simplicial complex* with the abstract complex. A set of points in a real space \mathbb{R}^D is *affinely independent* if it is not contained in a hyperplane; an affinely independent set in \mathbb{R}^D contains at most $D+1$ points. A geometric k -simplex is the *convex hull* of a set of $k+1$ affinely independent points, called its vertices. Hence, a point is a 0-simplex, a line segment is a 1-simplex, a triangle is a 2-simplex, a tetrahedron is a 3-simplex, and so on. A geometric simplicial complex shares the fundamental property of an abstract simplicial complex: It is a collection of simplices that is closed under inclusion and with the further property that the intersection of any two simplices in \mathcal{X} is also a simplex in \mathcal{X} , assuming that the empty set is an element of every simplicial complex. Although geometric simplicial complexes are easier to visualize and, for this reason, we will often use geometric terms like edges, triangles, and so on, as synonyms of pairs, triplets, we do not require the simplicial complex to be embedded in any real space, so as to leave the treatment as general as possible.

The structure of a simplicial complex is captured by the neighborhood relations of its subsets. As with graphs, it is useful to introduce first the orientation of the simplices. Every simplex, of any order, can have only two orientations, depending on the permutations of its elements. Two orientations are equivalent, or coherent, if each of them can be recovered from the other by an even number of transpositions, where

each transposition is a permutation of two elements [10]. A k -simplex $\sigma_i^k \equiv \{v_{i_0}, v_{i_1}, \dots, v_{i_k}\}$ of order k , together with an orientation is an *oriented* k -simplex and is denoted by $[v_{i_0}, v_{i_1}, \dots, v_{i_k}]$. Two simplices of order k , $\sigma_i^k, \sigma_j^k \in \mathcal{X}$, are *upper adjacent* in \mathcal{X} , if both are faces of a simplex of order $k+1$. Two simplices of order k , $\sigma_i^k, \sigma_j^k \in \mathcal{X}$, are *lower adjacent* in \mathcal{X} , if both have a common face of order $k-1$ in \mathcal{X} . A $(k-1)$ -face σ_j^{k-1} of a k -simplex σ_i^k is called a boundary element of σ_i^k . We use the notation $\sigma_j^{k-1} \subset \sigma_i^k$ to indicate that σ_j^{k-1} is a boundary element of σ_i^k . Given a simplex $\sigma_j^{k-1} \subset \sigma_i^k$, we use the notation $\sigma_j^{k-1} \sim \sigma_i^k$ to indicate that the orientation of σ_j^{k-1} is coherent with that of σ_i^k , whereas we write $\sigma_j^{k-1} \approx \sigma_i^k$ to indicate that the two orientations are opposite.

For each k , $C_k(\mathcal{X}, \mathbb{R})$ denotes the vector space obtained by the linear combination, using real coefficients, of the set of oriented k -simplices of \mathcal{X} . In algebraic topology, the elements of $C_k(\mathcal{X}, \mathbb{R})$ are called *k-chains*. If $\{\sigma_1^k, \dots, \sigma_{n_k}^k\}$ is the set of k -simplices in \mathcal{X} , a k -chain τ_k can be written as $\tau_k = \sum_{i=1}^{n_k} \alpha_i \sigma_i^k$. Then, given the basis $\{\sigma_1^k, \dots, \sigma_{n_k}^k\}$, a chain τ_k can be represented by the vector of its expansion coefficients $(\alpha_1, \dots, \alpha_{n_k})$. An important operator acting on ordered chains is the *boundary operator*. The boundary of the ordered k -chain $[v_{i_0}, \dots, v_{i_k}]$ is a linear mapping $\partial_k : C_k(\mathcal{X}, \mathbb{R}) \rightarrow C_{k-1}(\mathcal{X}, \mathbb{R})$ defined as

$$\partial_k[v_{i_0}, \dots, v_{i_k}] \triangleq \sum_{j=0}^k (-1)^j [v_{i_0}, \dots, v_{i_{j-1}}, v_{i_{j+1}}, \dots, v_{i_k}]. \quad (1)$$

So, for example, given an oriented triangle $\sigma_i^2 \triangleq [v_{i_0}, v_{i_1}, v_{i_2}]$, its boundary is

$$\partial_2 \sigma_i^2 = [v_{i_1}, v_{i_2}] - [v_{i_0}, v_{i_2}] + [v_{i_0}, v_{i_1}], \quad (2)$$

i.e., a suitable linear combination of its edges. It is straightforward to verify, by simple substitution, that *the boundary of a boundary is zero*, i.e., $\partial_k \partial_{k+1} = 0$.

It is important to remark that an oriented simplex is different from a directed one. As with graphs, an oriented edge establishes which direction of the flow is considered positive or negative, whereas a directed edge only permits flow in one direction [42]. In this work we will consider oriented, undirected simplices.

B. Algebraic representation

The structure of a simplicial complex \mathcal{X} of dimension K , shortly named K -simplicial complex, is fully described by the set of its incidence matrices \mathbf{B}_k , $k = 1, \dots, K$. Given an orientation of the simplicial complex \mathcal{X} , the entries of the incidence matrix \mathbf{B}_k establish which k -simplices are incident to which $(k-1)$ -simplices. Then \mathbf{B}_k is the matrix representation of the boundary operator. Formally speaking, its entries are defined as follows:

$$B_k(i, j) = \begin{cases} 0, & \text{if } \sigma_i^{k-1} \not\subset \sigma_j^k \\ 1, & \text{if } \sigma_i^{k-1} \subset \sigma_j^k \text{ and } \sigma_i^{k-1} \sim \sigma_j^k \\ -1, & \text{if } \sigma_i^{k-1} \subset \sigma_j^k \text{ and } \sigma_i^{k-1} \approx \sigma_j^k \end{cases}. \quad (3)$$

If we consider, for simplicity, a simplicial complex of order two, composed of a set \mathcal{V} of vertices, a set \mathcal{E} of edges, and

a set \mathcal{T} of triangles, having cardinalities $V = |\mathcal{V}|$, $E = |\mathcal{E}|$, and $T = |\mathcal{T}|$, respectively, we need to build two incidence matrices $\mathbf{B}_1 \in \mathbb{R}^{V \times E}$ and $\mathbf{B}_2 \in \mathbb{R}^{E \times T}$.

From (1), the property that the boundary of a boundary is zero maps into the following matrix form

$$\mathbf{B}_k \mathbf{B}_{k+1} = \mathbf{0}. \quad (4)$$

The structure of a K -simplicial complex is fully described by its *high order combinatorial Laplacian* matrices [43], of order $k = 0, \dots, K$, defined as

$$\mathbf{L}_0 = \mathbf{B}_1 \mathbf{B}_1^T, \quad (5)$$

$$\mathbf{L}_k = \mathbf{B}_k^T \mathbf{B}_k + \mathbf{B}_{k+1} \mathbf{B}_{k+1}^T, \quad k = 1, \dots, K-1, \quad (6)$$

$$\mathbf{L}_K = \mathbf{B}_K^T \mathbf{B}_K. \quad (7)$$

It is worth emphasizing that all Laplacian matrices of intermediate order, i.e. $k = 1, \dots, K-1$, contain two terms: The first term, also known as *lower Laplacian*, expresses the lower adjacency of k -order simplices; the second terms, also known as *upper Laplacian*, expresses the upper adjacency of k -order simplices. So, for example, two edges are lower adjacent if they share a common vertex, whereas they are upper adjacent if they are faces of a common triangle. Note that the vertices of a graph can only be upper adjacent, if they are incident to the same edge. This is why the graph Laplacian \mathbf{L}_0 contains only one term.

III. SPECTRAL SIMPLICIAL THEORY

In this paper, we are interested in analyzing signals defined over a simplicial complex. Given a set \mathcal{S}_k , with elements of cardinality $k+1$, a signal is defined as a real-valued map on the elements of \mathcal{S}_k , of the form

$$f_k : \mathcal{S}_k \rightarrow \mathbb{R}, \quad k = 0, 1, \dots \quad (8)$$

The order of the signal is one less the cardinality of the elements of \mathcal{S}_k . Even though our framework is general, in many cases we focus on simplices of order up to two. In that case, we consider a set of vertices \mathcal{V} , a set of edges \mathcal{E} and a set of triangles \mathcal{T} , of dimension V , E , and T , respectively. We denote with $\mathcal{X}(\mathcal{V}, \mathcal{E}, \mathcal{T})$ the associated simplicial complex. The signals over each complex of order k , with $k = 0, 1$ and 2 , are defined as the following maps: $\mathbf{s}^0 : \mathcal{V} \rightarrow \mathbb{R}^V$, $\mathbf{s}^1 : \mathcal{E} \rightarrow \mathbb{R}^E$, and $\mathbf{s}^2 : \mathcal{T} \rightarrow \mathbb{R}^T$.

Spectral graph theory represents a solid framework to extract features of a graph looking at the eigenvectors of the combinatorial Laplacian \mathbf{L}_0 of order 0. The eigenvectors associated with the smallest eigenvalues of \mathbf{L}_0 are very useful, for example, to identify clusters [44]. Furthermore, in GSP it is well known that a suitable basis to represent signals defined over the vertices of a graph, i.e. signals of order 0, is given by the eigenvectors of \mathbf{L}_0 . In particular, given the eigendecomposition of \mathbf{L}_0 :

$$\mathbf{L}_0 = \mathbf{U}_0 \mathbf{\Lambda}_0 \mathbf{U}_0^T, \quad (9)$$

the Graph Fourier Transform (GFT) of a signal \mathbf{s}^0 over an undirected graph has been defined as the projection of the

signal onto the space spanned by the eigenvectors of \mathbf{L}_0 , i.e. (see [3] and the references therein)

$$\widehat{\mathbf{s}}^0 \triangleq \mathbf{U}_0^T \mathbf{s}^0. \quad (10)$$

Equivalently, a signal defined over the vertices of a graph can be represented as

$$\mathbf{s}^0 = \mathbf{U}_0 \widehat{\mathbf{s}}^0. \quad (11)$$

From graph spectral theory, it is well known that the eigenvectors associated with the smallest eigenvalues of \mathbf{L}_0 encode information about the clusters of the graph. Hence, the representation given by (11) is particularly suitable for signals that are smooth within each cluster, whereas they can vary arbitrarily across different clusters. For such signals, in fact, the representation in (11) is *sparse* or approximately sparse.

As a generalization of the above approach, we may represent signals of various order over bases built with the eigenvectors of the corresponding high order Laplacian matrices, given in (6)-(7). Hence, using the eigendecomposition

$$\mathbf{L}_k = \mathbf{U}_k \mathbf{\Lambda}_k \mathbf{U}_k^T, \quad (12)$$

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

$$\widehat{\mathbf{s}}^k \triangleq \mathbf{U}_k^T \mathbf{s}^k, \quad (13)$$

so that a signal \mathbf{s}^k can be represented in terms of its GFT coefficients as

$$\mathbf{s}^k = \mathbf{U}_k \widehat{\mathbf{s}}^k. \quad (14)$$

Now we want to show under what conditions (14) is a meaningful representation of a k -order signal and what is the meaning of such a representation. More specifically, the goal of this section is threefold: i) we recall first the relations between eigenvectors of various order of \mathbf{L}_k ; ii) we recall the Hodge decomposition, which is a basic theory showing that the eigenvectors of any order can be split into three different classes, each representing a specific behavior of the signal; iii) we provide a theory showing how to build a *topology-aware* unitary basis to represent signals of various order starting only from topological properties.

A. Relations between eigenvectors of different order

There are interesting relations between the eigenvectors of Laplacian matrices of different order [45], [46] which is useful to recall as they play a key role in spectral analysis. The following properties hold true for the eigendecomposition of k -order Laplacian matrices, with $k = 1, \dots, K-1$.

Proposition 1: Given the Laplacian matrices \mathbf{L}_k of any order k , with $k = 1, \dots, K-1$, it holds:

- 1) the eigenvectors associated with the nonzero eigenvalues of $\mathbf{B}_k^T \mathbf{B}_k$ are orthogonal to the eigenvectors associated with the nonzero eigenvalues of $\mathbf{B}_{k+1}^T \mathbf{B}_{k+1}$ and vice-versa;
- 2) if \mathbf{v} is an eigenvector of $\mathbf{B}_k \mathbf{B}_k^T$ associated with the eigenvalue λ , then $\mathbf{B}_k^T \mathbf{v}$ is an eigenvector of $\mathbf{B}_k^T \mathbf{B}_k$, associated with the same eigenvalue;
- 3) the eigenvectors associated with the nonzero eigenvalues λ of \mathbf{L}_k are either the eigenvectors of $\mathbf{B}_k^T \mathbf{B}_k$ or those of $\mathbf{B}_{k+1}^T \mathbf{B}_{k+1}$;

- 4) the nonzero eigenvalues of \mathbf{L}_k are either the eigenvalues of $\mathbf{B}_k^T \mathbf{B}_k$ or those of $\mathbf{B}_{k+1}^T \mathbf{B}_{k+1}$.

Proof. All above properties are easy to prove. Property 1) is straightforward: If $\mathbf{B}_k^T \mathbf{B}_k \mathbf{v} = \lambda \mathbf{v}$, then

$$\mathbf{B}_{k+1} \mathbf{B}_{k+1}^T \lambda \mathbf{v} = \mathbf{B}_{k+1} \mathbf{B}_{k+1}^T \mathbf{B}_k^T \mathbf{B}_k \mathbf{v} = \mathbf{0} \quad (15)$$

because of (4). Similarly, for the converse. Property 2) is also straightforward: If \mathbf{v} is an eigenvector of $\mathbf{B}_k \mathbf{B}_k^T$ associated with a nonvanishing eigenvalue λ , then

$$(\mathbf{B}_k^T \mathbf{B}_k) \mathbf{B}_k^T \mathbf{v} = \mathbf{B}_k^T (\mathbf{B}_k \mathbf{B}_k^T) \mathbf{v} = \lambda \mathbf{B}_k^T \mathbf{v}. \quad (16)$$

Finally, properties 3) and 4) follow from the definition of k -order Laplacian, i.e. $\mathbf{L}_k = \mathbf{B}_k^T \mathbf{B}_k + \mathbf{B}_{k+1} \mathbf{B}_{k+1}^T$ and from property 1). ■

Remark: Recalling that the eigenvectors associated with the smallest nonzero eigenvalues of \mathbf{L}_0 are smooth within each cluster, applying property 2) to the case $k = 1$, it turns out that the eigenvectors of \mathbf{L}_1 associated with the smallest eigenvalues of $\mathbf{B}_1^T \mathbf{B}_1$ are approximately null over the links within each cluster, whereas they assume the largest (in modulus) values on the edges across clusters. These eigenvectors are then useful to highlight *inter-cluster edges*.

B. Hodge decomposition

Let us consider the eigendecomposition of the k -th order Laplacian, for $k = 1, \dots, K-1$,

$$\mathbf{L}_k = \mathbf{B}_k^T \mathbf{B}_k + \mathbf{B}_{k+1} \mathbf{B}_{k+1}^T = \mathbf{U}_k \mathbf{\Lambda}_k \mathbf{U}_k^T. \quad (17)$$

The structure of \mathbf{L}_k , together with the property $\mathbf{B}_k \mathbf{B}_{k+1} = \mathbf{0}$, induces an interesting decomposition of the space \mathbb{R}^{D_k} of signals of order k of dimension D_k . First of all, the property $\mathbf{B}_k \mathbf{B}_{k+1} = \mathbf{0}$ implies $\text{img}(\mathbf{B}_{k+1}) \subseteq \ker(\mathbf{B}_k)$. Hence, each vector $\mathbf{x} \in \ker(\mathbf{B}_k)$ can be decomposed into two parts: one belonging to $\text{img}(\mathbf{B}_{k+1})$ and one orthogonal to it. Furthermore, since the whole space \mathbb{R}^{D_k} can always be written as $\mathbb{R}^{D_k} \equiv \ker(\mathbf{B}_k) \oplus \text{img}(\mathbf{B}_k^T)$, it is possible to decompose \mathbb{R}^{D_k} into the direct sum [47]

$$\mathbb{R}^{D_k} \equiv \text{img}(\mathbf{B}_k^T) \oplus \ker(\mathbf{L}_k) \oplus \text{img}(\mathbf{B}_{k+1}), \quad (18)$$

where the vectors in $\ker(\mathbf{L}_k)$ are also in $\ker(\mathbf{B}_k)$ and $\ker(\mathbf{B}_{k+1}^T)$. This implies that, given any signal \mathbf{s}^k of order k , there always exist three signals \mathbf{s}^{k-1} , \mathbf{s}_H^k , and \mathbf{s}^{k+1} , of order $k-1$, k , and $k+1$, respectively, such that \mathbf{s}^k can always be expressed as the sum of three *orthogonal* components:

$$\mathbf{s}^k = \mathbf{B}_k^T \mathbf{s}^{k-1} + \mathbf{s}_H^k + \mathbf{B}_{k+1} \mathbf{s}^{k+1}. \quad (19)$$

This decomposition is known as *Hodge decomposition* [48] and it is the extension of the Hodge theory for differential forms on Riemannian manifold to simplicial complexes. The subspace $\ker(\mathbf{L}_k)$ is called harmonic subspace since each $\mathbf{s}_H^k \in \ker(\mathbf{L}_k)$ is a solution of the *discrete* Laplace equation

$$\mathbf{L}_k \mathbf{s}_H^k = (\mathbf{B}_k^T \mathbf{B}_k + \mathbf{B}_{k+1} \mathbf{B}_{k+1}^T) \mathbf{s}_H^k = \mathbf{0}.$$

When embedded in a real space, a fundamental property of geometric simplicial complexes of order k is that the dimensions of $\ker(\mathbf{L}_k)$, for $k = 0, \dots, K$ are *topological invariants*

of the K -simplicial complex, i.e. topological features that are preserved under homeomorphic transformations of the space. The dimensions of $\ker(\mathbf{L}_k)$ are also known as *Betti numbers* β_k of order k : β_0 is the number of connected components of the graph, β_1 is the number of holes, β_2 is the number of cavities, and so on [48].

The decomposition in (19) shows an interesting interplay between signals of different order, which we will exploit in the ensuing sections. Before proceeding, it is useful to clarify the meaning of the terms appearing in (19). Let us consider, for simplicity, the analysis of flow signals, i.e. the case $k = 1$. To this end, it is useful to introduce the curl and divergence operators, in analogy with their continuous time counterpart operators applied to vector fields. More specifically, given an edge signal \mathbf{s}^1 , the (discrete) curl operator is defined as

$$\text{curl}(\mathbf{s}^1) = \mathbf{B}_2^T \mathbf{s}^1. \quad (20)$$

This operator maps the edge signal \mathbf{s}^1 onto a signal defined over the triangle sets, i.e. in \mathbb{R}^T , and it is straightforward to verify that the generic i -th entry of the $\text{curl}(\mathbf{s}^1)$ is a measure of the *flow circulating along the edges of the i -th triangle*. As an example for the simplicial complex in Fig. 1, we have

$$\mathbf{B}_2^T = \begin{bmatrix} 0 & 0 & 1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 1 & 0 & 0 \end{bmatrix} \quad (21)$$

and, defining $\mathbf{s}^1 = [e_1, e_2, \dots, e_{10}, e_{11}]^T$, we get $\text{curl}(\mathbf{s}^1) = [e_3 - e_4 + e_5, e_2 - e_3 + e_7, -e_7 + e_8 + e_9]^T$. Each entry of \mathbf{s}^1 is then the circulation over the corresponding triangle.

Similarly, the (discrete) divergence operator maps the edge signal \mathbf{s}^1 onto a signal defined over the vertex space, i.e. \mathbb{R}^V , and it is defined as

$$\text{div}(\mathbf{s}^1) = \mathbf{B}_1 \mathbf{s}^1. \quad (22)$$

Again, by direct substitution, it turns out that the i -th entry of $\text{div}(\mathbf{s}^1)$ represents the *net-flow passing through the i -th vertex*, i.e. the difference between the inflow and outflow at node i . Thus a non-zero divergence reveals the presence of a source or sink node. For the example in Fig. 1, we get

$$\mathbf{B}_1 = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & -1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (23)$$

so that $\text{div}(\mathbf{s}^1) = [-e_1 - e_6, e_1 - e_2 - e_3 - e_4, e_2 - e_7 - e_9 - e_{10}, e_{10} - e_{11}, -e_8 + e_9 + e_{11}, e_3 - e_5 + e_7 + e_8, e_4 + e_5 + e_6]^T$. If we consider equation (19) in the case $k = 1$,

$$\mathbf{s}^1 = \mathbf{B}_1^T \mathbf{s}^0 + \mathbf{s}_H^1 + \mathbf{B}_2 \mathbf{s}^2, \quad (24)$$

recalling that $\mathbf{B}_1 \mathbf{B}_2 = \mathbf{0}$, it is easy to check that the first component in (24) has zero curl, and then it may be called an *irrotational* component, whereas the third component has zero divergence, and then it may be called a *solenoidal* component, in analogy to the calculus terminology used for vector fields. The *harmonic* component \mathbf{s}_H^1 is a flow vector that is both

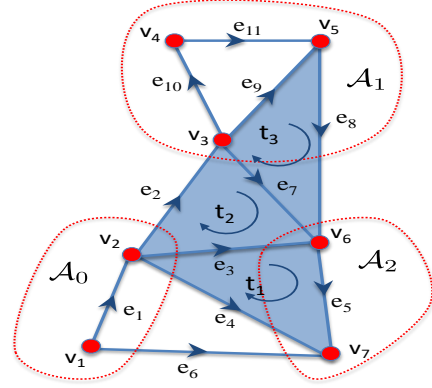


Fig. 1: Cut of order 1.

curl-free and divergence-free. Notice also that, in (24), $\mathbf{B}_1^T \mathbf{s}^0$ represents the (discrete) *gradient* of \mathbf{s}^0 .

C. Topology-aware unitary basis

In this section, we propose a method to build a unitary basis to represent edge signals, reflecting topological properties of the complex, more specifically triple-wise relations. The idea underlying the method is to search a basis that minimizes a real-valued function capturing triple-wise relations. For example, in the graph case, a key topological property is connectivity, which is well captured by the *cut* function. The associated real-valued function can be built using the so called *Lovász extension* [49], [50] of the cut size. More specifically, given a graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ and a partition of its vertex set \mathcal{V} in two sets \mathcal{A}_0 and \mathcal{A}_1 , with $\mathcal{A}_0 \cup \mathcal{A}_1 = \mathcal{V}$ and $\mathcal{A}_0 \cap \mathcal{A}_1 = \emptyset$, the cut size is defined as

$$F_0(\mathcal{A}_0, \mathcal{A}_1) = \text{cut}(\mathcal{A}_0, \mathcal{A}_1) \triangleq \sum_{i \in \mathcal{A}_0} \sum_{j \in \mathcal{A}_1} a_{ij} \quad (25)$$

where $a_{ij} = 1$ if $(i, j) \in \mathcal{E}$ and $a_{ij} = 0$ otherwise. $F_0(\mathcal{A}_0, \mathcal{A}_1)$ is a *set* function and is known to be a submodular function [49]. Now, we want to translate the set function $F_0(\mathcal{A}_0, \mathcal{A}_1)$ onto a real-valued function $f_0(\mathbf{x}^0)$, defined over \mathbb{R}^V , to be used for the subsequent optimization. This can be done using the so called *Lovász extension* [49], which is equal to:

$$f_0(\mathbf{x}^0) = \sum_{i=1}^V \sum_{j=1}^V a_{ij} |x_i^0 - x_j^0| \quad (26)$$

with $\mathbf{x}^0 \in \mathbb{R}^V$. The function in (26) measures the total variation of a signal defined over the nodes of a graph and then it can be used as a regularization function, whenever the signal of interest is known to be a smooth function. The function in (26) formed the basis of the method proposed in [51] to build a unitary basis for analyzing signals defined over the vertices of a graph. More specifically, in [51] the basis was built by solving the following optimization problem

$$\mathbf{U} \triangleq (\mathbf{u}_1, \dots, \mathbf{u}_V) = \arg \min_{\mathbf{U} \in \mathbb{R}^{V \times V}} \sum_{n=2}^V f_0(\mathbf{u}_n) \quad \text{s.t.} \quad \mathbf{U}^T \mathbf{U} = \mathbf{I}, \quad \mathbf{u}_1 = \frac{1}{\sqrt{V}} \mathbf{1} \quad (27)$$

where $\mathbf{1}$ is the $V \times 1$ vector of all ones. In the above problem, the objective function to be minimized is convex, but the problem is non-convex because of the unitarity constraint. To simplify the search for the basis, the objective function in (27) can be relaxed to become

$$f_0^r(\mathbf{x}^0) = \sum_{i=1}^V \sum_{j=1}^V a_{ij} (x_i^0 - x_j^0)^2. \quad (28)$$

Substituting (28) in (27), we still have a non-convex problem. However, its solution is known to be given by the eigenvectors of \mathbf{L}_0 . From this perspective, the basis typically used in the GFT, built as the eigenvectors of the Laplacian matrix, can be interpreted as the solution of the above optimization problem, after relaxation.

Now, we extend the previous approach to find a regularization function capturing triple-wise relations, useful to analyze signals defined over the *edges* of a graph. In the previous case, the analysis of *node* signals started from the *bi-partition* of a graph and then the construction of a real-valued extension of the cut size. Here, to analyze *edge* signals, we need to start from the *tri-partition* of the discrete set and define a set function counting the number of *triangles* gluing the three sets together. The function to be minimized should then be the real valued (Lovász) extension of such a set function.

The combinatorial study of simplicial complexes has attracted increasing attention and the generalization of Cheeger inequalities to simplicial complexes has been the subject of several works [52], [53]. In particular, Hein *et al.* introduced the total variation on hypergraphs as the Lovász extension of the hypergraph cut, defined as the size of the hyperedges set connecting a bipartition of the vertex set [54]. In this work, we derive a Lovász extension *defined over the edge set*. More formally, we study a simplicial complex $\mathcal{X}(\mathcal{V}, \mathcal{E}, \mathcal{T})$ of order two, as in the example sketched in Fig.1, and consider the partition of \mathcal{V} in *three* sets $(\mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2)$. The *triple-wise coupling* function is now defined as

$$F_1(\mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2) = \sum_{i \in \mathcal{A}_0} \sum_{j \in \mathcal{A}_1} \sum_{k \in \mathcal{A}_2} a_{ijk} \quad (29)$$

where $a_{ijk} = 1$ if $(i, j, k) \in \mathcal{T}$ and $a_{ijk} = 0$ otherwise. Our main result, stated in the following theorem, is that the Lovász extension of the triple-wise coupling function gives rise to a measure of the curl of edge signals along triangles.

Theorem 1: *Let $\mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2$ be a partition of the vertex set \mathcal{V} of the 2-dimensional simplicial complex $\mathcal{X}(\mathcal{V}, \mathcal{E}, \mathcal{T})$ with $|\mathcal{E}| = E$, $|\mathcal{T}| = T$. Then the Lovász extension $f : \mathbb{R}^E \rightarrow \mathbb{R}$, evaluated at $\mathbf{x}^1 \in \mathbb{R}^E$, of the triple-wise coupling size $F_1(\mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2)$ defined in (29), is*

$$f_1(\mathbf{x}^1) = \sum_{i,j,k=1}^E a_{ijk} |x_i^1 - x_j^1 + x_k^1| = \sum_{j=1}^T \left| \sum_{i=1}^E B_2(i, j) x_i^1 \right| \quad (30)$$

where $B_2(i, j)$ are the *edge-triplet incidence coefficients* defined in (3).

Proof. Please see Appendix A. ■

The function in (30) represents the sum of the absolute values of the curls over all the triangles. Differently from [54],

where the total variation over a hypergraph was built from a bipartition of a discrete set and it was a function defined over \mathbb{R}^V , in our case, we start from a *tripartition* of the original set and we build a real-valued extension, defined over \mathbb{R}^E , i.e. a space of dimension equal to the number of edges. A suitable unitary basis for representing (curling) edge signals can then be found as the solution of the following optimization problem

$$\begin{aligned} \mathbf{U} \triangleq (\mathbf{u}_1, \dots, \mathbf{u}_E) &= \arg \min_{\mathbf{U} \in \mathbb{R}^{E \times E}} \sum_{n=1}^E f_1(\mathbf{u}_n) \\ \text{s.t.} \quad &\mathbf{U}^T \mathbf{U} = \mathbf{I}. \end{aligned} \quad (31)$$

The objective function is convex, but the problem is non-convex because of the unitarity constraint. The above problem can be solved resorting to the algorithm proposed in [51], tuned to the new objective function given in (30). Alternatively, similarly to what is typically done with graphs, $f_1(\mathbf{x}^1)$ can be relaxed and replaced with

$$f_1^r(\mathbf{x}^1) = \sum_{j=1}^T \left(\sum_{i=1}^E B_2(i, j) x_i^1 \right)^2 = \mathbf{x}^{1T} \mathbf{B}_2 \mathbf{B}_2^T \mathbf{x}^1. \quad (32)$$

Substituting this function back to (31), the solution is given by the eigenvectors of $\mathbf{B}_2 \mathbf{B}_2^T$.

The above arguments show that the eigenvectors of $\mathbf{B}_2 \mathbf{B}_2^T$ provide a suitable basis to analyze edge signals having a curling behavior. However, as we know from the Hodge decomposition recalled in the previous section, edge signals contain other useful components that are orthogonal to solenoidal signals, namely irrotational and harmonic components. Then, in general, it is useful to take as a unitary basis for analyzing edge signals *all* the eigenvectors of \mathbf{L}_1 , i.e. the eigenvectors associated to the nonzero eigenvalues of $\mathbf{B}_2 \mathbf{B}_2^T$ and of $\mathbf{B}_1^T \mathbf{B}_1$, plus the eigenvectors associated to the kernel of \mathbf{L}_1 . In summary, the theory developed in this section provides a further argument, rooted on intrinsic topological properties of the simplicial complex on which the signal is defined, to exploit the Hodge decomposition to find a suitable basis for the analysis of edge signals. Furthermore, the theory says that using the Laplacian eigenvectors comes as a consequence of a relaxation step. In many cases, whenever the numerical complexity is not an issue, it may be better to keep the ℓ_1 -norm objective functions given in (26) and (30), as this would yield more sparse, and then more informative, representations, as a generalization of what done for directed graphs in [51].

IV. EDGE FLOWS ESTIMATION

Let us consider now the observation of an edge signal affected by additive noise. Our goal in this section, is to formulate the denoising problem as a constrained problem, rooted on the Hodge decomposition. Denoising edge signals embedded in noise was already considered in [42], [31], [32] and, more specifically, in [18]. The formulation proposed in [18] was based on the definition of signals over simplicial complexes as skew-symmetric arrays of dimension equal to the order of the corresponding simplex plus one. So, the edge flow was represented as a matrix (i.e., a two-dimensional array) satisfying the constraint $X(i, j) = -X(j, i)$. A signal defined over

the triangles was represented as an array of dimension three, satisfying the property $\Phi(i, j, k) = \Phi(j, k, i) = \Phi(k, i, j) = -\Phi(j, i, k) = -\Phi(i, k, j) = -\Phi(k, j, i)$. Our aim in this section, is to formulate the denoising optimization problem dealing only with vectors. Based on (19), the observed vector can be modeled as

$$\mathbf{x}^1 = \mathbf{B}_1^T \mathbf{s}^0 + \mathbf{s}_H^1 + \mathbf{B}_2 \mathbf{s}^2 + \mathbf{v}^1 \quad (33)$$

where \mathbf{v}^1 is noise. Let us suppose, for simplicity, that the noise vector is Gaussian, with zero-mean entries all having the same variance σ_n^2 . The optimal estimator can then be formulated as the solution of the following problem

$$\begin{aligned} (\hat{\mathbf{s}}^0, \hat{\mathbf{s}}^2, \hat{\mathbf{s}}_H^1) = & \underset{\mathbf{s}^0 \in \mathbb{R}^V, \mathbf{s}^2 \in \mathbb{R}^T, \mathbf{s}_H^1 \in \mathbb{R}^E}{\operatorname{argmin}} \quad \|\mathbf{B}_2 \mathbf{s}^2 + \mathbf{B}_1^T \mathbf{s}^0 + \mathbf{s}_H^1 - \mathbf{x}^1\|^2 \\ \text{s.t.} \quad & \mathbf{B}_1 \mathbf{s}_H^1 = \mathbf{0} \\ & \mathbf{B}_2^T \mathbf{s}_H^1 = \mathbf{0} \quad (\mathcal{Q}). \end{aligned} \quad (34)$$

Note that problem \mathcal{Q} is convex. Then, there exists multipliers $\lambda_1 \in \mathbb{R}^V, \lambda_2 \in \mathbb{R}^T$ such that the tuple $(\hat{\mathbf{s}}^0, \hat{\mathbf{s}}^2, \hat{\mathbf{s}}_H^1, \lambda_1, \lambda_2)$ satisfies the Karush-Kuhn-Tucker (KKT) conditions of \mathcal{Q} (note that Slater's constraint qualification is satisfied [55]). The associated Lagrangian function is

$$\begin{aligned} \mathcal{L}(\mathbf{s}^0, \mathbf{s}^2, \mathbf{s}_H^1, \lambda_1, \lambda_2) = & (\mathbf{B}_2 \mathbf{s}^2 + \mathbf{B}_1^T \mathbf{s}^0 + \mathbf{s}_H^1 - \mathbf{x}^1)^T \\ & \cdot (\mathbf{B}_2 \mathbf{s}^2 + \mathbf{B}_1^T \mathbf{s}^0 + \mathbf{s}_H^1 - \mathbf{x}^1) + \lambda_1^T \mathbf{B}_1 \mathbf{s}_H^1 + \lambda_2^T \mathbf{B}_2^T \mathbf{s}_H^1. \end{aligned} \quad (35)$$

Exploiting the orthogonality property $\mathbf{B}_1 \mathbf{B}_2 = \mathbf{0}$, it is easy to get the following KKT conditions

- (a) $\nabla_{\mathbf{s}^0} \mathcal{L}(\mathbf{s}^0, \mathbf{s}^2, \mathbf{s}_H^1, \lambda_1, \lambda_2) = \mathbf{B}_1 \mathbf{B}_1^T \mathbf{s}^0 - \mathbf{B}_1 \mathbf{x}^1 = \mathbf{0}$
- (b) $\nabla_{\mathbf{s}^2} \mathcal{L}(\mathbf{s}^0, \mathbf{s}^2, \mathbf{s}_H^1, \lambda_1, \lambda_2) = \mathbf{B}_2^T \mathbf{B}_2 \mathbf{s}^2 - \mathbf{B}_2^T \mathbf{x}^1 = \mathbf{0}$
- (c) $\nabla_{\mathbf{s}_H^1} \mathcal{L}(\mathbf{s}^0, \mathbf{s}^2, \mathbf{s}_H^1, \lambda_1, \lambda_2) = \mathbf{s}_H^1 - \mathbf{x}^1 + \mathbf{B}_1^T \lambda_1 + \mathbf{B}_2 \lambda_2 = \mathbf{0}$
- (d) $\mathbf{B}_1 \mathbf{s}_H^1 = \mathbf{0}, \mathbf{B}_2^T \mathbf{s}_H^1 = \mathbf{0}$
- (e) $\lambda_1 \in \mathbb{R}^V, \lambda_2 \in \mathbb{R}^T$.

Note that conditions (a)-(c) reduce to

$$\begin{aligned} \text{(a)} \quad & \mathbf{L}_0 \mathbf{s}^0 = \mathbf{B}_1 \mathbf{x}^1 \\ \text{(b)} \quad & \mathbf{B}_2^T \mathbf{B}_2 \mathbf{s}^2 = \mathbf{B}_2^T \mathbf{x}^1 \\ \text{(c)} \quad & \mathbf{s}_H^1 = \mathbf{x}^1 - \mathbf{B}_1^T \lambda_1 - \mathbf{B}_2 \lambda_2. \end{aligned} \quad (36)$$

Multiplying both sides of condition (c) by \mathbf{B}_2^T , and using the second equality in (d) and condition (b), we get

$$\mathbf{B}_2^T \mathbf{B}_2 \mathbf{s}^2 = \mathbf{B}_2^T \mathbf{B}_2 \lambda_2. \quad (37)$$

This means that \mathbf{s}^2 and λ_2 may differ only by an additive vector lying in the nullspace of $\mathbf{B}_2^T \mathbf{B}_2$. Let us set $\mathbf{s}^2 = \lambda_2 + \mathbf{c}_2$, with $\mathbf{B}_2^T \mathbf{B}_2 \mathbf{c}_2 = \mathbf{0}$. Similarly, multiplying (c) by \mathbf{B}_1 and using the first equality in (d) and condition (a), we obtain

$$\mathbf{B}_1 \mathbf{B}_1^T \mathbf{s}^0 = \mathbf{B}_1 \mathbf{B}_1^T \lambda_1, \quad (38)$$

which implies that $\mathbf{s}^0 = \lambda_1 + \mathbf{c}_1$, with \mathbf{c}_1 such that $\mathbf{B}_1 \mathbf{B}_1^T \mathbf{c}_1 = \mathbf{0}$. Thus, condition (c) reduces to

$$\mathbf{s}_H^1 = \mathbf{x}^1 - \mathbf{B}_1^T \mathbf{s}^0 - \mathbf{B}_2 \mathbf{s}^2 \quad (39)$$

which says, as expected, that we can derive the harmonic component by subtracting the estimated solenoidal and irrotational parts from the observed flow signal \mathbf{x}^1 . To recover the

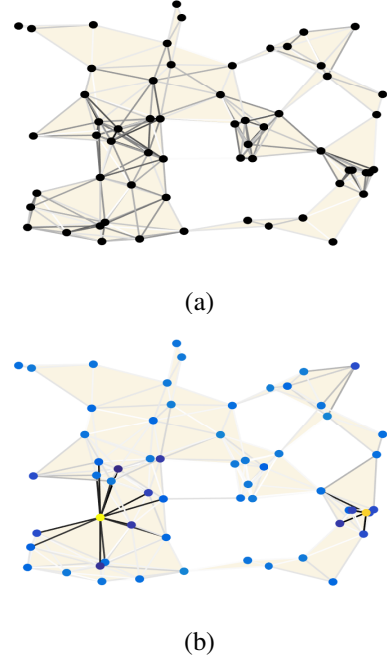


Fig. 2: Observed flow on a simplicial complex (a) and reconstruction of the irrotational flow (b).

irrotational flow $\mathbf{s}_{\text{irr}}^1$ from the 0-order signal \mathbf{s}^0 we need to solve equation (a) in (36). Note that \mathbf{L}_0 is not invertible. For connected graphs, it has rank $V-1$ and its kernel is the span of the vector $\mathbf{1}$ of all ones. However, since the vector $\mathbf{b} = \mathbf{B}_1 \mathbf{x}^1$ is also orthogonal to $\mathbf{1}$, the normal equation $\mathbf{L}_0 \mathbf{s}^0 = \mathbf{B}_1 \mathbf{x}^1$ admits the nontrivial solution (at least for connected graphs):

$$\mathbf{s}^0 = \mathbf{L}_0^\dagger \mathbf{B}_1 \mathbf{x}^1$$

where \dagger denoted the Moore-Penrose pseudo-inverse.

Similarly, the 2-order signal \mathbf{s}^2 , solution of the second equation in (36), can be obtained as

$$\mathbf{s}^2 = (\mathbf{B}_2^T \mathbf{B}_2)^\dagger \mathbf{B}_2^T \mathbf{x}^1$$

since $\mathbf{B}_2^T \mathbf{x}^1$ is orthogonal to the null space of $\mathbf{B}_2^T \mathbf{B}_2$. The irrotational, solenoidal and harmonic components can then be recovered as follows

$$\begin{aligned} \hat{\mathbf{s}}_{\text{irr}}^1 &= \mathbf{B}_1^T \hat{\mathbf{s}}^0 = \mathbf{B}_1^T \mathbf{L}_0^\dagger \mathbf{B}_1 \mathbf{x}^1 \\ \hat{\mathbf{s}}_{\text{sol}}^1 &= \mathbf{B}_2 \hat{\mathbf{s}}^2 = \mathbf{B}_2 (\mathbf{B}_2^T \mathbf{B}_2)^\dagger \mathbf{B}_2^T \mathbf{x}^1 \\ \hat{\mathbf{s}}_H^1 &= \mathbf{x}^1 - \hat{\mathbf{s}}_{\text{irr}}^1 - \hat{\mathbf{s}}_{\text{sol}}^1. \end{aligned} \quad (40)$$

Note that the first two conditions in (36) imply that the variables $\mathbf{s}^0, \mathbf{s}^2$ in \mathcal{Q} are indeed decoupled so that the optimal solutions coincide with those of the following problems

$$\hat{\mathbf{s}}^0 = \underset{\mathbf{s}^0 \in \mathbb{R}^V}{\operatorname{argmin}} \quad \|\mathbf{B}_1^T \mathbf{s}^0 - \mathbf{x}^1\|^2 \quad (\mathcal{Q}_0) \quad (41)$$

$$\hat{\mathbf{s}}^2 = \underset{\mathbf{s}^2 \in \mathbb{R}^T}{\operatorname{argmin}} \quad \|\mathbf{B}_2 \mathbf{s}^2 - \mathbf{x}^1\|^2 \quad (\mathcal{Q}_2). \quad (42)$$

An example of an edge (flow) signal is reported in Fig. 2(a), representing the simulation of data packet flow over a network, including measurement errors. The signal values are encoded in the gray color associated to each link. Suppose now, that

the goal of processing is to detect nodes injecting anomalous traffic in the network, starting from the traffic shown in Fig. 2(a). If some node is a source of an anomalous traffic, that node generates an edge signal with a strong irrotational (or divergence-like) component. To detect spamming nodes, we can then project the overall observed traffic vector \mathbf{s}^1 onto the space orthogonal to the space spanned by the nonzero eigenvectors of $\mathbf{B}_2\mathbf{B}_2^T$, computing

$$\mathbf{y}^1 = (\mathbf{I} - \mathbf{U}_{\text{sol}}\mathbf{U}_{\text{sol}}^T) \mathbf{s}^1 \quad (43)$$

where \mathbf{U}_{sol} is the matrix whose columns collect the eigenvectors associated with the nonzero eigenvalues of $\mathbf{B}_2\mathbf{B}_2^T$. The result of this projection is reported in Fig. 2(b), where we can clearly see that the only edges with a significant contributions are the ones located around two nodes, i.e. the source nodes injecting traffic in the network, whose divergence is encoded by the node color.

V. SAMPLING AND RECOVERING OF SIGNAL DEFINED OVER SIMPLICIAL COMPLEX

Suppose now that we only observe a few samples of a k -order signal. The question we address here is to find the conditions to recover the whole signal \mathbf{s}^k from a subset of samples. To answer this question, we may use the theory developed in [56] for signals on graph, and later extended to hypergraphs in [57]. For simplicity, we focus on signals defined over a simplicial complex of order $K = 2$, i.e. on vertices, edges and triangles. Given a set of edges $\mathcal{S} \subseteq \mathcal{E}$ we define an edge-limiting operator as a diagonal matrix $\mathbf{D}_{\mathcal{S}}$ of dimension equal to the number of edges, with a one in the positions where we measure the flow, and zero elsewhere, i.e.

$$\mathbf{D}_{\mathcal{S}} = \text{diag}\{\mathbf{1}_{\mathcal{S}}\} \quad (44)$$

where $\mathbf{1}_{\mathcal{S}}$ is the set indicator vector whose i -th entry is equal to one if the edge $e_i \in \mathcal{S}$ and zero otherwise. We say that an edge signal \mathbf{s}^1 is perfectly localized over the subset $\mathcal{S} \subseteq \mathcal{E}$ (or \mathcal{S} -edge-limited) if $\mathbf{s}^1 = \mathbf{D}_{\mathcal{S}}\mathbf{s}^1$. Similarly, given the matrix \mathbf{U}_1 whose columns are the eigenvectors of \mathbf{L}_1 , and a subset of indices \mathcal{F} , we define the operator

$$\mathbf{F}_{\mathcal{F}} = \mathbf{U}_1 \mathbf{\Sigma}_{\mathcal{F}} \mathbf{U}_1^T \quad (45)$$

where $\mathbf{\Sigma}_{\mathcal{F}} = \text{diag}(\mathbf{1}_{\mathcal{F}})$. An edge signal \mathbf{s}^1 is \mathcal{F} -bandlimited over a frequency set \mathcal{F} if $\mathbf{F}_{\mathcal{F}}\mathbf{s}^1 = \mathbf{s}^1$. The operators $\mathbf{D}_{\mathcal{S}}$ and $\mathbf{F}_{\mathcal{F}}$ are self-adjoint and idempotent and represents orthogonal projectors, respectively, on the sets \mathcal{S} and \mathcal{F} . If we look for edges signals which are perfectly localized in both the edge and frequency domains, some conditions for perfect localization have been derived in [56]. More specifically: a) \mathbf{s}^1 is perfectly localized over both the edge set \mathcal{S} and the frequency set \mathcal{F} if and only if the operator $\mathbf{F}_{\mathcal{F}}\mathbf{D}_{\mathcal{S}}\mathbf{F}_{\mathcal{F}}$ has an eigenvalue equal to one, i.e.

$$\|\mathbf{D}_{\mathcal{S}}\mathbf{F}_{\mathcal{F}}\|_2 = \|\mathbf{F}_{\mathcal{F}}\mathbf{D}_{\mathcal{S}}\|_2 = \|\mathbf{F}_{\mathcal{F}}\mathbf{D}_{\mathcal{S}}\mathbf{F}_{\mathcal{F}}\|_2 = 1 \quad (46)$$

where $\|\mathbf{A}\|_2$ denotes the spectral norm of \mathbf{A} ; b) a sufficient condition for the existence of such signals is that $|\mathcal{S}| + |\mathcal{F}| > E$. Conversely, if $|\mathcal{S}| + |\mathcal{F}| \leq E$, there could still exist perfectly localized vectors, when condition (46) holds.

In the following we first consider sampling on a single layer by extracting samples from the edge signals and then, we consider multi-layer processing using samples of signals defined over different layers of the simplex.

A. Single-layer sampling

In the following theorem [56] we provide a necessary and sufficient condition to recover the edge signal \mathbf{s}^1 from its samples $\mathbf{s}_{\mathcal{S}}^1 \triangleq \mathbf{D}_{\mathcal{S}}\mathbf{s}^1$.

Theorem 2: *Given the bandlimited edge signal $\mathbf{s}^1 = \mathbf{F}_{\mathcal{F}}\mathbf{s}^1$, it is possible to recover \mathbf{s}^1 from a subset of samples collected over the subset $\mathcal{S} \subseteq \mathcal{E}$ if and only if the following condition holds:*

$$\|\bar{\mathbf{D}}_{\mathcal{S}}\mathbf{F}_{\mathcal{F}}\|_2 = \|\mathbf{F}_{\mathcal{F}}\bar{\mathbf{D}}_{\mathcal{S}}\|_2 < 1 \quad (47)$$

with $\bar{\mathbf{D}}_{\mathcal{S}} = \mathbf{I} - \mathbf{D}_{\mathcal{S}}$.

Proof. The proof is a straightforward extension of Th. 4.1 in [56] to signals defined on the edges of the complex. ■

In words, the above conditions mean that there can be no \mathcal{F} -bandlimited signals that are perfectly localized on the complementary set $\bar{\mathcal{S}}$. Perfect recovery of the signal \mathbf{s}^1 from $\mathbf{s}_{\mathcal{S}}^1$ can be achieved as

$$\mathbf{r}^1 = \mathbf{Q}_{\mathcal{S}}\mathbf{s}_{\mathcal{S}}^1 \quad (48)$$

where $\mathbf{Q}_{\mathcal{S}} = (\mathbf{I} - \bar{\mathbf{D}}_{\mathcal{S}}\mathbf{F}_{\mathcal{F}})^{-1}$. The existence of the above inverse is ensured by condition (47). In fact, the reconstruction error can be written as [56]

$$\mathbf{s}^1 - \mathbf{Q}_{\mathcal{S}}\mathbf{s}_{\mathcal{S}}^1 = \mathbf{s}^1 - \mathbf{Q}_{\mathcal{S}}(\mathbf{I} - \bar{\mathbf{D}}_{\mathcal{S}})\mathbf{s}^1 = \mathbf{s}^1 - \mathbf{Q}_{\mathcal{S}}(\mathbf{I} - \bar{\mathbf{D}}_{\mathcal{S}}\mathbf{F}_{\mathcal{F}})\mathbf{s}^1 = \mathbf{0},$$

where we exploited in the second equality the bandlimited condition $\mathbf{s}^1 = \mathbf{F}_{\mathcal{F}}\mathbf{s}^1$.

To make (47) holds true, we must guarantee that $\mathbf{D}_{\mathcal{S}}\mathbf{F}_{\mathcal{F}}\mathbf{s}^1 \neq \mathbf{0}$ or, equivalently, that the matrix $\mathbf{D}_{\mathcal{S}}\mathbf{F}_{\mathcal{F}}$ is full column rank, i.e. $\text{rank}(\mathbf{D}_{\mathcal{S}}\mathbf{F}_{\mathcal{F}}) = |\mathcal{F}|$. Then, a necessary condition to ensure this holds is $|\mathcal{S}| \geq |\mathcal{F}|$.

An alternative way to retrieve the overall signal \mathbf{s}^1 from its samples can be obtained as follows. If (47) holds true, the entire signal \mathbf{s}^1 can be recovered from $\mathbf{s}_{\mathcal{S}}^1$ as follows

$$\mathbf{s}^1 = \mathbf{U}_{\mathcal{F}} (\mathbf{U}_{\mathcal{F}}^T \mathbf{D}_{\mathcal{S}} \mathbf{U}_{\mathcal{F}})^{-1} \mathbf{U}_{\mathcal{F}}^T \mathbf{s}_{\mathcal{S}}^1 \quad (49)$$

where $\mathbf{U}_{\mathcal{F}}$ is the $E \times |\mathcal{F}|$ matrix whose columns are the eigenvectors of \mathbf{L}_1 associated with the frequency set \mathcal{F} .

Remark: It is worth to notice that, because of the Hodge decomposition (18), an edge signal always contains three components that are typically band-limited, as they reside on a subspace of dimension smaller than E . This means that, if one knows *a priori*, that the edge signal contains only one component, e.g. solenoidal, irrotational, or harmonic, then it is possible to observe the edge signal and to recover the desired component over all the edges, under the conditions established by Theorem 2.

B. Multi-layer sampling

In this section, we consider the case where we take samples of signals of *different order* and we propose two alternative strategies to retrieve an edge signal \mathbf{s}^1 from these samples. The first approach aims at recovering \mathbf{s}^1 by using both, the vertex signal samples $\mathbf{s}_{\mathcal{A}}^0 = \mathbf{D}_{\mathcal{A}}\mathbf{s}^0$, with $\mathcal{A} \subseteq \mathcal{V}$, and the edge

samples $\mathbf{s}_S^1 = \mathbf{D}_S \mathbf{s}^1$. Hereinafter, we denote by \mathcal{F}_{irr} , \mathcal{F}_{sol} and \mathcal{F}_H the set of frequency indexes in \mathcal{F} corresponding to the eigenvectors of \mathbf{L}_1 belonging, respectively, to the irrotational, solenoidal and harmonic subspaces. Note that, if \mathbf{s}^1 is \mathcal{F} -bandlimited then also \mathbf{s}_H^1 , $\mathbf{s}_{\text{sol}}^1$ and $\mathbf{s}_{\text{irr}}^1$ are bandlimited with bandwidth, respectively, $|\mathcal{F}_H|$, $|\mathcal{F}_{\text{sol}}|$ and $|\mathcal{F}_{\text{irr}}|$. Define \mathcal{F}_{SH} as the set of frequency indexes such that $\mathcal{F}_{\text{SH}} = \mathcal{F}_H \cup \mathcal{F}_{\text{sol}}$. Furthermore, given the matrix \mathbf{U}_0 with columns the eigenvectors of \mathbf{L}_0 , we define the operator $\mathbf{F}_{\mathcal{F}_0}^0 = \mathbf{U}_0 \Sigma_{\mathcal{F}_0} \mathbf{U}_0^T$ where $|\mathcal{F}_0|$ denotes the bandwidth of \mathbf{s}^0 . Let \mathcal{C}_1 be the set of indexes in \mathcal{F}_0 associated with the eigenvectors belonging to $\ker(\mathbf{L}_0)$ and denote by $\mathbf{U}_{\mathcal{F}_0 - \mathcal{C}_1}^0$ the matrix whose columns are the eigenvectors of \mathbf{L}_0 associated with the frequency set $\mathcal{F}_0 - \mathcal{C}_1$. Then, we can state the following theorem.

Theorem 3: Consider the second-order simplex $\mathcal{X}(\mathcal{V}, \mathcal{E}, \mathcal{T})$ and the edge signal $\mathbf{s}^1 = \mathbf{s}_{\text{sol}}^1 + \mathbf{s}_H^1 + \mathbf{B}_1^T \mathbf{s}^0$. Then, assume that: i) the vertex-signal \mathbf{s}^0 and the edge signal \mathbf{s}^1 are bandlimited with bandwidth, respectively, $|\mathcal{F}_0|$ and $|\mathcal{F}| = |\mathcal{F}_{\text{SH}}| + |\mathcal{F}_0| - c_1$, where $|\mathcal{F}_{\text{SH}}| = |\mathcal{F}_{\text{sol}}| + |\mathcal{F}_H|$ and $c_1 \geq 0$ denotes the number of eigenvectors in the bandwidth of \mathbf{s}^0 belonging to $\ker(\mathbf{L}_0)$; ii) the conditions $\|\bar{\mathbf{D}}_{\mathcal{A}} \mathbf{F}_{\mathcal{F}_0}^0\|_2 < 1$ and $\|\bar{\mathbf{D}}_S \mathbf{F}_{\mathcal{F}_{\text{SH}}}\|_2 < 1$ hold true. Then, it follows that:

- a) \mathbf{s}^1 can be perfectly recovered from both a set of vertex signal samples $\mathbf{s}_{\mathcal{A}}^0 = \mathbf{D}_{\mathcal{A}} \mathbf{s}^0$ and from the edge samples $\mathbf{s}_S^1 = \mathbf{D}_S \mathbf{s}^1$ via the set of equations

$$\begin{bmatrix} \mathbf{s}^0 \\ \bar{\mathbf{s}}^1 \end{bmatrix} = \mathbf{Q} \begin{bmatrix} \mathbf{s}_{\mathcal{A}}^0 \\ \mathbf{s}_S^1 \end{bmatrix}, \quad (50)$$

where $\bar{\mathbf{s}}^1 = \mathbf{s}_{\text{sol}}^1 + \mathbf{s}_H^1$,

$$\mathbf{Q} = \begin{bmatrix} (\mathbf{I} - \bar{\mathbf{D}}_{\mathcal{A}} \mathbf{F}_{\mathcal{F}_0}^0)^{-1} & \mathbf{O} \\ \mathbf{P} & (\mathbf{I} - \bar{\mathbf{D}}_S \mathbf{F}_{\mathcal{F}_{\text{SH}}})^{-1} \end{bmatrix} \quad (51)$$

and $\mathbf{P} = -(\mathbf{I} - \bar{\mathbf{D}}_S \mathbf{F}_{\mathcal{F}_{\text{SH}}})^{-1} \mathbf{D}_S \mathbf{B}_1^T \mathbf{F}_{\mathcal{F}_0}^0 (\mathbf{I} - \bar{\mathbf{D}}_{\mathcal{A}} \mathbf{F}_{\mathcal{F}_0}^0)^{-1}$;

- b) There exists a set of frequency indexes $\mathcal{F}_{\text{irr}} \subset \mathcal{F}$, for which the eigenvectors matrix $\mathbf{U}_{\mathcal{F}_{\text{irr}}}$ of \mathbf{L}_1 satisfies the equality $\mathbf{U}_{\mathcal{F}_{\text{irr}}} = \mathbf{B}_1^T \mathbf{U}_{\mathcal{F}_0 - \mathcal{C}_1}^0$, and such that any \mathcal{F} -bandlimited edge signal with set of frequency indexes $\mathcal{F} = \mathcal{F}_{\text{irr}} \cup \mathcal{F}_{\text{SH}}$ and $|\mathcal{F}_{\text{irr}}| = |\mathcal{F}_0| - c_1$ can be recovered by using $N_0 \geq |\mathcal{F}_0|$ samples from \mathbf{s}^0 and $N_1 \geq |\mathcal{F}_{\text{SH}}|$ samples from \mathbf{s}^1 .

Proof. See the supporting material in [58, Appendix B]. ■

Let us now consider the case where we want to recover \mathbf{s}^1 from samples collected from signals of order 0, 1, and 2, i.e. \mathbf{s}^0 , \mathbf{s}^1 and \mathbf{s}^2 . We denote by $\mathbf{s}_{\mathcal{M}}^2 = \mathbf{D}_{\mathcal{M}} \mathbf{s}^2$ the vector of triangle signal samples with $\mathcal{M} \subseteq \mathcal{T}$. Furthermore, given the matrix \mathbf{U}_2 with columns the eigenvectors of the second-order Laplacian \mathbf{L}_2 , we define the operator $\mathbf{F}_{\mathcal{F}_2}^2 = \mathbf{U}_2 \Sigma_{\mathcal{F}_2} \mathbf{U}_2^T$ where $|\mathcal{F}_2|$ denotes the bandwidth of \mathbf{s}^2 . Then, assuming \mathbf{s}^2 bandlimited, it holds $\mathbf{s}^2 = \mathbf{F}_{\mathcal{F}_2}^2 \mathbf{s}^2$. Denote with \mathcal{C}_2 the set of indexes in \mathcal{F}_2 associated with the eigenvectors belonging to $\ker(\mathbf{L}_2)$ and with $\mathbf{U}_{\mathcal{F}_2 - \mathcal{C}_2}^2$ the matrix whose columns are the eigenvectors of \mathbf{L}_2 associated with the index set $\mathcal{F}_2 - \mathcal{C}_2$.

Theorem 4: Consider the second-order simplex $\mathcal{X}(\mathcal{V}, \mathcal{E}, \mathcal{T})$, the edge signal $\mathbf{s}^1 = \mathbf{B}_2 \mathbf{s}^2 + \mathbf{s}_H^1 + \mathbf{B}_1^T \mathbf{s}^0$ and assume that: i) the vertex-signal \mathbf{s}^0 , the edge signal \mathbf{s}^1 and the triangle signal \mathbf{s}^2 are bandlimited with bandwidth,

respectively, $|\mathcal{F}_0|$, $|\mathcal{F}_1| = |\mathcal{F}_0| + |\mathcal{F}_H| + |\mathcal{F}_2| - (c_1 + c_2)$ and $|\mathcal{F}_2|$, with $c_1 \geq 0$ and $c_2 \geq 0$ the number of eigenvectors in the bandwidth of \mathbf{s}^0 and \mathbf{s}^2 belonging, respectively, to $\ker(\mathbf{L}_0)$ and $\ker(\mathbf{L}_2)$; ii) all conditions $\|\bar{\mathbf{D}}_{\mathcal{A}} \mathbf{F}_{\mathcal{F}_0}^0\|_2 < 1$, $\|\bar{\mathbf{D}}_S \mathbf{F}_{\mathcal{F}_H}\|_2 < 1$ and $\|\bar{\mathbf{D}}_{\mathcal{M}} \mathbf{F}_{\mathcal{F}_2}^2\|_2 < 1$ hold true. Then, it follows:

- a) \mathbf{s}^1 can be perfectly recovered from a set of vertex signal samples $\mathbf{s}_{\mathcal{A}}^0 = \mathbf{D}_{\mathcal{A}} \mathbf{s}^0$, from the edge samples $\mathbf{s}_S^1 = \mathbf{D}_S \mathbf{s}^1$ and from the triangle samples $\mathbf{s}_{\mathcal{M}}^2 = \mathbf{D}_{\mathcal{M}} \mathbf{s}^2$ as

$$\begin{bmatrix} \mathbf{s}^0 \\ \mathbf{s}_H^1 \\ \mathbf{s}^2 \end{bmatrix} = \mathbf{R} \begin{bmatrix} \mathbf{s}_{\mathcal{A}}^0 \\ \mathbf{s}_S^1 \\ \mathbf{s}_{\mathcal{M}}^2 \end{bmatrix}, \quad (52)$$

where

$$\mathbf{R} = \begin{bmatrix} (\mathbf{I} - \bar{\mathbf{D}}_{\mathcal{A}} \mathbf{F}_{\mathcal{F}_0}^0)^{-1} & \mathbf{O} & \mathbf{O} \\ \mathbf{P}_1 & \mathbf{P}_2 & \mathbf{P}_3 \\ \mathbf{O} & \mathbf{O} & (\mathbf{I} - \bar{\mathbf{D}}_{\mathcal{M}} \mathbf{F}_{\mathcal{F}_2}^2)^{-1} \end{bmatrix} \quad (53)$$

and $\mathbf{P}_1 = -(\mathbf{I} - \bar{\mathbf{D}}_S \mathbf{F}_{\mathcal{F}_H})^{-1} \mathbf{D}_S \mathbf{B}_1^T \mathbf{F}_{\mathcal{F}_0}^0 (\mathbf{I} - \bar{\mathbf{D}}_{\mathcal{A}} \mathbf{F}_{\mathcal{F}_0}^0)^{-1}$, $\mathbf{P}_2 = (\mathbf{I} - \bar{\mathbf{D}}_S \mathbf{F}_{\mathcal{F}_H})^{-1}$, $\mathbf{P}_3 = -(\mathbf{I} - \bar{\mathbf{D}}_S \mathbf{F}_{\mathcal{F}_H})^{-1} \mathbf{D}_S \mathbf{B}_2 \mathbf{F}_{\mathcal{F}_2}^2 (\mathbf{I} - \bar{\mathbf{D}}_{\mathcal{M}} \mathbf{F}_{\mathcal{F}_2}^2)^{-1}$;

- b) There exist two sets of frequency indexes $\mathcal{F}_{\text{sol}}, \mathcal{F}_{\text{irr}} \subset \mathcal{F}$, for which the eigenvectors of \mathbf{L}_1 stacked in the columns of the matrices $\mathbf{U}_{\mathcal{F}_{\text{sol}}}, \mathbf{U}_{\mathcal{F}_{\text{irr}}}$ satisfy, respectively, the equality $\mathbf{U}_{\mathcal{F}_{\text{sol}}} = \mathbf{B}_2 \mathbf{U}_{\mathcal{F}_2 - \mathcal{C}_2}^2$, and $\mathbf{U}_{\mathcal{F}_{\text{irr}}} = \mathbf{B}_1^T \mathbf{U}_{\mathcal{F}_0 - \mathcal{C}_1}^0$. Then, any \mathcal{F} -bandlimited edge signal with frequency set $\mathcal{F} = \mathcal{F}_{\text{sol}} \cup \mathcal{F}_{\text{irr}} \cup \mathcal{F}_H$, and $|\mathcal{F}_{\text{irr}}| = |\mathcal{F}_0| - c_1$, $|\mathcal{F}_{\text{sol}}| = |\mathcal{F}_2| - c_2$, can be recovered by using $N_0 \geq |\mathcal{F}_0|$ samples from \mathbf{s}^0 , $N_1 \geq |\mathcal{F}_H|$ samples from \mathbf{s}^1 and $N_2 \geq |\mathcal{F}_2|$ samples from \mathbf{s}^2 .

Proof. See the supporting material in [58, Appendix B]. ■

VI. ESTIMATION OF DISCRETE VECTOR FIELDS

Developing tools to process signals defined over simplicial complexes is also useful to devise algorithms for processing discrete vector fields. The use of algebraic topology, and more specifically *Discrete Exterior Calculus* (DEC), has been already considered for the analysis of vector fields, especially in the field of computer graphic. Exterior Calculus (EC) is a discipline that generalizes vector calculus to smooth manifolds of arbitrary dimensions [59] and DEC is a discretization of EC on simplicial complexes [60], [61]. DEC methodologies have been already proposed in [62] to produce smooth tangent vector fields in computer graphics. Smoothing tangential vector fields using techniques based on the spectral decomposition of higher order Laplacian was also advocated in [63]. In this section, building on the basic ideas of [62], [63], we propose an alternative approach to smooth discrete vector fields. More specifically, as in [62], [63], the proposed approach is based on three main steps: i) map the vector field onto an edge signal; ii) filter the edge signal; and iii) map the edge signal back to a vector field. Steps i) and iii) are essentially the same as in [63]. The difference we introduce here is in step ii), i.e. in the filtering of the edge signal. Filtering edge signals has been considered before, see, e.g., [32], but in our case we consider a

different formulation and we incorporate the proper metrics in the Hodge Laplacian, dictated from the initial mapping from the vector field to the corresponding edge signal.

To filter vector fields, we need first to embed the vector field and the simplicial complex into a real space. Let us denote with \mathcal{X} a simplicial complex embedded in a real space \mathbb{R}^n of dimension n . We assume that \mathcal{X} is flat, i.e. all simplices are in the same affine n -subspace, and well-centered, which means the circumcenter of each simplex lies in its interior [60]. A discrete vector field \vec{v} can be considered as a map from points \mathbf{x}_i of \mathcal{X} to \mathbb{R}^n . An illustrative example is shown in Fig. 3(a), where the discrete vector field is represented by the set of arrows associated to the points in a regular grid in \mathbb{R}^2 . In many applications, it is of interest to develop tools to extract relevant information from a vector field or to reduce the effect of noise. Our proposed approach is based on the following three main steps:

1) *Map a vector field onto a simplicial complex signal*

Given the set of N points \mathbf{x}_i , we build the simplicial complex \mathcal{X} using a well-centered Delaunay triangulation with vertices (0-simplices) in the points \mathbf{x}_i [60]. Then, starting from the set of vectors $\vec{v}(\mathbf{x}_i)$, we recover a continuous vector field through interpolation based on the Whitney basis functions $\phi_i(\mathbf{x})$ [64]

$$\vec{v}(\mathbf{x}) = \sum_{i=1}^N \vec{v}(\mathbf{x}_i) \phi_i(\mathbf{x}) \quad (54)$$

where $\phi_i(\mathbf{x})$ are affine piecewise functions. More specifically, ϕ_i is an hat function on vertex v_i , which takes on the value one at vertex v_i , i.e. $\phi_i(\mathbf{x}_i) = 1$, is zero at all other vertices, and affine over each 2-simplex having v_i as vertex. Then, project the vector field onto the set of 1-simplices (edges), giving rise to a set of scalar signals defined over the edges of the complex

$$x_j^1 = \left(\frac{\vec{v}(\sigma_{j_0}^0) + \vec{v}(\sigma_{j_1}^0)}{2} \right) \cdot \vec{\sigma}_j^1, \quad j = 1, \dots, E \quad (55)$$

where j_0 and j_1 are the end-points of edge j and $\vec{\sigma}_j^1$ stands for the vector corresponding to $\sigma_j^1 = [\sigma_{j_0}^0, \sigma_{j_1}^0]$, having the same direction as the orientation of σ_j^1 . The vector $\mathbf{x}^1 \in \mathbb{R}^E$ of edge signals associated with the vector field \vec{v} , is then built as $\mathbf{x}^1 = [x_1^1, \dots, x_E^1]^T$.

2) *Filter the signal defined on the simplicial complex*

The filtering strategy aims to recover an edge flow vector \mathbf{s}^1 that fits the observed vector \mathbf{x}^1 , it is smooth and sparse. We recover the vector \mathbf{s}^1 as the solution of

$$\min_{\mathbf{s}^1 \in \mathbb{R}^E} \|\mathbf{s}^1 - \mathbf{x}^1\|_2 + \lambda \mathbf{s}^{1T} \mathbf{L}_1 \mathbf{s}^1 + \gamma \|\mathbf{s}^1\|_1 \quad (\mathcal{P}_{\mathcal{F}})$$

where λ, γ are positive penalty coefficients controlling, respectively, the signal smoothness and its sparsity. Since we chose the Whitney form as interpolation basis, for any two signals of order p , $\mathbf{x}^p, \mathbf{y}^p \in \mathbb{R}^{N_p}$, the induced inner product is given by $\mathbf{x}^{pT} \mathbf{M}_p \mathbf{y}^p$, where the $N_p \times N_p$ -dimensional matrix \mathbf{M}_p , incorporating the Whitney metric, is derived as in [61][Prop. 9.7]. Therefore, the Hodge Laplacian \mathbf{L}_1 , weighted with the appropriate metric matrices \mathbf{M}_p , with $p = 0, 1, 2$, is a semidefinite positive matrix that can be written as

$$\mathbf{L}_1 = \mathbf{B}_2 \mathbf{M}_2 \mathbf{B}_2^T + \mathbf{M}_1 \mathbf{B}_1^T \mathbf{M}_0^{-1} \mathbf{B}_1 \mathbf{M}_1 \quad (56)$$

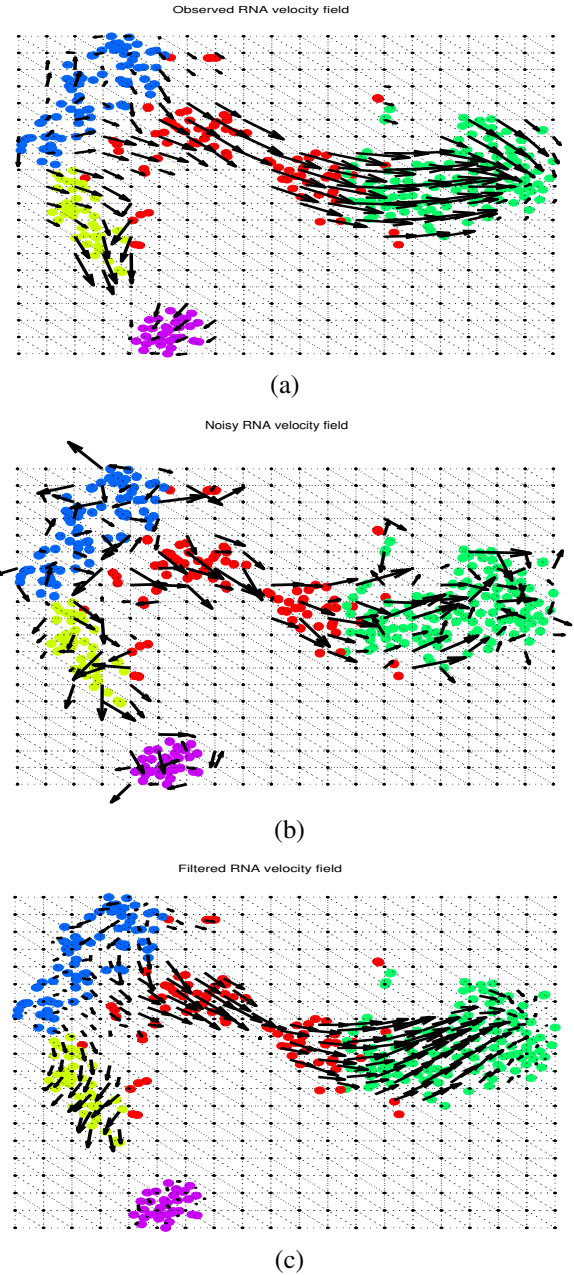


Fig. 3: RNA velocity fields: (a) observed and (b) noisy fields; (c) reconstructed field.

while the fitting error norm becomes

$$\|\mathbf{s}^1 - \mathbf{x}^1\|_2 = (\mathbf{s}^1 - \mathbf{x}^1)^T \mathbf{M}_1 (\mathbf{s}^1 - \mathbf{x}^1). \quad (57)$$

3) *Map the filtered signal back onto a discrete vector field*

Finally, given the discrete signal of order 1, $s_{e_{ij}}^1, \forall e_{ij} \in \mathcal{E}$, the generated piecewise linear vector field becomes [62]

$$\vec{v}(\mathbf{x}) = \sum_{e_{ij}} s_{e_{ij}}^1 [\phi_i(\mathbf{x}) \nabla \phi_j(\mathbf{x}) - \phi_j(\mathbf{x}) \nabla \phi_i(\mathbf{x})]. \quad (58)$$

We show now an interesting application of the above procedure to the analysis of the vector field representing the RNA velocity field, defined as the time derivative of the gene expression state [41], useful to predict the future state

of individual cells and then the direction of change of the entire transcriptome during the RNA sequencing dynamic process. The RNA velocity can be estimated by distinguishing between nascent (unspliced) and mature (spliced) messenger RNA (mRNA) in common single-cell RNA sequencing protocols [41]. We consider the mRNA-seq dataset of mouse chromaffin cell differentiation analysed in [41]. An example of RNA velocity vector field is illustrated in Fig. 3(a). To analyze such a vector field using the proposed algorithms, we implemented steps 1) and 3) using the discrete exterior calculus operators provided by the PyDEC software developed in [61]. We considered a Delaunay well-centered triangulation of the continuous bi-dimensional space, which generates the simplicial complexes in Fig. 3 composed of $N = 400$ nodes, where we fill all the triangles. The velocity field in Fig. 3(a) is observed over 156 vertices and the field vector at each vertex has been obtained with a local Gaussian kernel smoothing [41]. The underlying colored cells represented different states of the cell differentiation process. Then, to test our filtering strategy, we added a strong Gaussian noise to the observed velocity field, with SNR = 0 dB, as illustrated in Fig. 3(b). This noise is added to incorporate mRNA molecules degradation and model inaccuracy. Then we apply the proposed filtering strategy by first reconstructing the edge signal as in (54) and then recovering the edge signal as a solution of the optimization problem $\mathcal{P}_{\mathcal{F}}$. Finally, we reconstruct the vector field using the interpolation formula (58), observed at the barycentric points of each triangles. The result is reported in Fig. 3(c), where we can appreciate the robustness of the proposed filtering strategy.

VII. INFERENCE OF SIMPLICIAL COMPLEX TOPOLOGY FROM DATA

The inference of the graph topology from (node) signals is a problem that has received significant attention, as shown in the recent tutorial papers [65], [39], [40] and in the references therein. In this section, we propose algorithms to infer the structure of a simplicial complex. Given the layer structure of a simplicial complex, we propose a *hierarchical* approach that infers the structure of one layer, assuming knowledge of the lower order layers. For simplicity, we focus on the inference of a complex of order 2 from the observation of a set of M edge (flow) signals $\mathbf{X}^1 := [\mathbf{x}^1(1), \dots, \mathbf{x}^1(M)]$, assuming that the topology of the underlying graph is given (or it has been estimated). So, we start from the knowledge of \mathbf{L}_0 , which implies, after selection of an orientation, knowledge of \mathbf{B}_1 . Since $\mathbf{L}_1 = \mathbf{B}_1^T \mathbf{B}_1 + \mathbf{B}_2 \mathbf{B}_2^T$, then we need to estimate \mathbf{B}_2 . Before doing that, we check, from the data, if the term $\mathbf{B}_2 \mathbf{B}_2^T$ is really needed. Since, from (24), the only components that may depend on \mathbf{B}_2 are the solenoidal and harmonic components, we first project the observed flow signal onto the space orthogonal to the space spanned by the irrotational component, by computing

$$\mathbf{x}_{\text{SH}}^1(m) = (\mathbf{I} - \mathbf{U}_{\text{irr}} \mathbf{U}_{\text{irr}}^T) \mathbf{x}^1(m), m = 1, \dots, M, \quad (59)$$

where \mathbf{U}_{irr} is the matrix whose columns are the eigenvectors associated with the nonzero eigenvalues of $\mathbf{B}_1^T \mathbf{B}_1$. Then, denoting with $\mathbf{X}_{\text{SH}}^1 = [\mathbf{x}_{\text{SH}}^1(1), \dots, \mathbf{x}_{\text{SH}}^1(M)]$ the signal matrix

of size $E \times M$, we measure the energy of \mathbf{X}_{SH}^1 by taking its norm $\|\mathbf{X}_{\text{SH}}^1\|_F$: If the norm is smaller than a threshold η of the averaged energy of the observed data set, we stop, otherwise we proceed to estimate \mathbf{B}_2 .

The first step in the estimation of \mathbf{B}_2 starts from the detection of all cliques of three elements present in the graph. Their number is $T = \text{trace}[(\mathbf{L}_0 - \text{diag}(\mathbf{L}_0 \mathbf{1}))^3] / 6$. For each clique, we choose, arbitrarily, an orientation for the potential triangle filling it. The matrix \mathbf{B}_2 can then be written as

$$\mathbf{B}_2 = \sum_{n=1}^T t_n \mathbf{b}_n \mathbf{b}_n^T \quad (60)$$

where \mathbf{b}_n is the vector of size E associated with the n -th clique, whose entries are all zero except the three entries associated with the three edges of the n -th clique. Those entries assume the value 1 or -1 , depending on the orientation of the triangle associated with the n -th clique. The coefficients t_n in (60) are equal to one, if there is a (filled) triangle on the n -th clique, or zero otherwise. The goal of our inference algorithm is then to decide, starting from the data, which entries of $\mathbf{t} := [t_1, \dots, t_T]$ are equal to one or zero. Our strategy is to make the association that enforces a small total variation of the observed flow signal on the inferred complex, using (32) as a measure of total variation on flow signals. We propose two alternative algorithms: The first method infers the structure of \mathbf{B}_2 by minimizing the total variation of the observed data; the second method performs first a Principal Component Analysis (PCA) and then looks for the matrix \mathbf{B}_2 and the coefficients of the expansion over the principal components that minimize the total variation plus a penalty on the model fitting error.

Minimum Total Variation (MTV) Algorithm: The goal of this algorithm is to minimize the total variation over the observed data set, assuming knowledge of the number of triangles. The set of coefficients \mathbf{t} is found as solution of

$$\begin{aligned} \min_{\mathbf{t} \in \{0,1\}^T} \quad & q(\mathbf{t}) \triangleq \sum_{n=1}^T t_n \text{trace} \left(\mathbf{X}_{\text{SH}}^{1T} \mathbf{b}_n \mathbf{b}_n^T \mathbf{X}_{\text{SH}}^1 \right) \quad (\mathcal{P}_{\text{MTV}}) \\ \text{s.t.} \quad & \|\mathbf{t}\|_0 = t^*, \quad t_n \in \{0,1\}, \forall n, \end{aligned} \quad (61)$$

where t^* is the number of triangles that we aim to detect. In practice, this number is not known, so it has to be found through cross-validation. Even though problem \mathcal{P}_{MTV} is non-convex, it can be solved in closed form. Introducing the nonnegative coefficients $c_n = \sum_{i=1}^M \mathbf{x}_{\text{SH}}^{1T}(i) \mathbf{b}_n \mathbf{b}_n^T \mathbf{x}_{\text{SH}}^1(i)$, the solution can in fact be obtained by sorting the coefficients c_n in increasing order and then selecting the triangles associated with the indices of the t^* lowest coefficients c_n . Note that the proposed strategy infers the presence of triangles along the cliques having the minimum curl along its edges. Hence, we expect better performance when the edge signal contains only the harmonic components, whose curls along the filled triangles is exactly null.

PCA-based Best Fitting with Minimum Total Variation (PCA-BFMTV): To robustify the MTV algorithm in the case where the edge signal contains also a solenoidal component and is possibly corrupted by noise, we propose now the PCA-

BFMTV algorithm that infers the structure of \mathbf{B}_2 and the edge signal that best fits the observed data set \mathbf{X}^1 , while at the same time exhibiting a small total variation over the inferred topology. The method starts performing a principal component analysis of the observed data by extracting the eigenvectors associated with the largest eigenvalues of the covariance matrix estimated from the observed data set. More specifically, the proposed strategy is composed of two steps: 1) estimate the covariance matrix $\hat{\mathbf{C}}_X$ from the edge signal data set \mathbf{X}_{SH}^1 and builds the matrix $\hat{\mathbf{U}}_{\text{SH}}$ whose columns are the eigenvectors associated with the F largest eigenvalues of $\hat{\mathbf{C}}_X$; 2) model the observed data set as $\mathbf{X}_{\text{SH}}^1 = \hat{\mathbf{U}}_{\text{SH}} \hat{\mathbf{S}}_{\text{SH}}^1$ and searches for the coefficient matrix $\hat{\mathbf{S}}_{\text{SH}}^1$ and the vector \mathbf{t} that solve the following problem

$$\begin{aligned} \min_{\mathbf{t} \in \{0,1\}^T, \hat{\mathbf{S}}_{\text{SH}}^1 \in \mathbb{R}^{F \times M}} \quad & g(\mathbf{t}, \hat{\mathbf{S}}_{\text{SH}}^1) + \gamma \|\mathbf{X}_{\text{SH}}^1 - \hat{\mathbf{U}}_{\text{SH}} \hat{\mathbf{S}}_{\text{SH}}^1\|_F^2 \\ \text{s.t.} \quad & \|\mathbf{t}\|_0 = t^*, \quad t_n \in \{0,1\}, \forall n, \quad (\mathcal{P}_{\mathcal{F}_S}) \end{aligned} \quad (62)$$

where $g(\mathbf{t}, \hat{\mathbf{S}}_{\text{SH}}^1) \triangleq \sum_{n=1}^T t_n \text{trace}(\hat{\mathbf{S}}_{\text{SH}}^1 T \hat{\mathbf{U}}_{\text{SH}}^T \mathbf{b}_n \mathbf{b}_n^T \hat{\mathbf{U}}_{\text{SH}} \hat{\mathbf{S}}_{\text{SH}}^1)$ and γ is a non-negative coefficient controlling the trade-off between the data fitting error and the signal smoothness. Although problem $\mathcal{P}_{\mathcal{F}_S}$ is non-convex, it can be solved using an iterative alternating optimization algorithm returning successive estimates of $\hat{\mathbf{S}}_{\text{SH}}^1$, having fixed \mathbf{t} , and alternately \mathbf{t} , given $\hat{\mathbf{S}}_{\text{SH}}^1$. Interestingly, each step of the alternating optimization problem admits a closed form solution. More specifically, at each iteration k , the coefficient matrix $\hat{\mathbf{S}}_{\text{SH}}^1[k]$ can be found as

$$\hat{\mathbf{S}}_{\text{SH}}^1[k] = \arg \min_{\hat{\mathbf{S}}_{\text{SH}}^1 \in \mathbb{R}^{F \times M}} g(\mathbf{t}[k], \hat{\mathbf{S}}_{\text{SH}}^1) + \gamma \|\mathbf{X}_{\text{SH}}^1 - \hat{\mathbf{U}}_{\text{SH}} \hat{\mathbf{S}}_{\text{SH}}^1\|_F^2 \quad (\mathcal{P}_{\mathcal{F}_S}^k).$$

Defining $\mathbf{L}_{\text{upp}}[k] := \sum_{n=1}^T t_n[k] \mathbf{b}_n \mathbf{b}_n^T$, problem $\mathcal{P}_{\mathcal{F}_S}^k$ admits the closed form solution

$$\hat{\mathbf{S}}_{\text{SH}}^1[k] = (\mathbf{I}_F + \gamma \hat{\mathbf{U}}_{\text{SH}}^T \mathbf{L}_{\text{upp}}[k] \hat{\mathbf{U}}_{\text{SH}})^{-1} \hat{\mathbf{U}}_{\text{SH}}^T \mathbf{X}_{\text{SH}}^1. \quad (63)$$

Then, given $\hat{\mathbf{S}}_{\text{SH}}^1[k]$, we can find the vector $\mathbf{t}[k+1]$ using the same method used to solve problem MTV, in (61), i.e. setting $c_n[k] := \text{trace}(\hat{\mathbf{S}}_{\text{SH}}^1[k]^T \hat{\mathbf{U}}_{\text{SH}}^T \mathbf{b}_n \mathbf{b}_n^T \hat{\mathbf{U}}_{\text{SH}} \hat{\mathbf{S}}_{\text{SH}}^1[k])$ and taking the entries of $\mathbf{t}[k+1]$ equal to 1 for the indices corresponding to the first t^* smallest coefficients of $\{c_n[k]\}_{n=1}^T$, and 0 otherwise. The iterative steps of the proposed strategy are reported in the box entitled Algorithm PCA-BFMTV. Now we test the validity of our inference algorithms over both simulated and real data.

Performance on synthetic data: Some of the most critical parameters affecting the goodness of the proposed algorithms are the dimension of the subspaces associated with the solenoidal and harmonic components of the signal and the number of filled triangles in the complex. In fact, in both MTV and PCA-BFMTV a key aspect is the detection of triangles as the cliques where the associated curl is minimum. Hence, if the signal contains only the harmonic component and there is no noise, the triangles can be identified with no error, because the harmonic component is null over the filled triangles.

Algorithm PCA-BFMTV

Set $\gamma > 0$, $\mathbf{t}[0] \in \{0,1\}^T$, $\|\mathbf{t}[0]\|_0 = t^*$,

$$\mathbf{L}_{\text{upp}}[0] = \sum_{n=1}^T t_n[0] \mathbf{b}_n \mathbf{b}_n^T, \quad k = 1$$

Repeat

$$\text{Set } \hat{\mathbf{S}}_{\text{SH}}^1[k] = (\mathbf{I}_F + \gamma \hat{\mathbf{U}}_{\text{SH}}^T \mathbf{L}_{\text{upp}}[k-1] \hat{\mathbf{U}}_{\text{SH}})^{-1} \hat{\mathbf{U}}_{\text{SH}}^T \mathbf{X}_{\text{SH}}^1;$$

Compute $\mathbf{t}[k]$ by sorting the coefficients

$$c_n[k] = \text{trace}(\hat{\mathbf{S}}_{\text{SH}}^1[k]^T \hat{\mathbf{U}}_{\text{SH}}^T \mathbf{b}_n \mathbf{b}_n^T \hat{\mathbf{U}}_{\text{SH}} \hat{\mathbf{S}}_{\text{SH}}^1[k]),$$

and setting to 1 the entries of $\mathbf{t}[k]$ corresponding to the t^* smallest coefficients, and 0 otherwise;

Set $k = k + 1$,

until convergence.

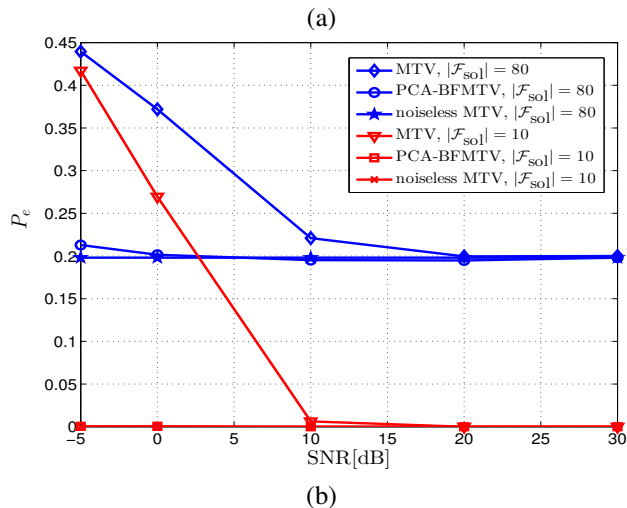
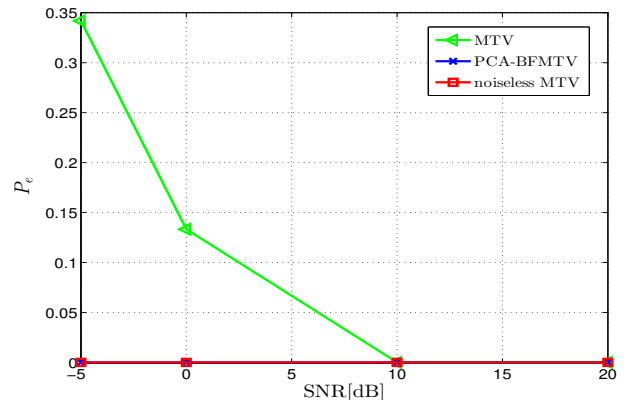


Fig. 4: Error probability by observing: (a) harmonic noisy signals; (b) harmonic plus solenoidal noisy signals.

However, when there is a solenoidal component or noise, there might be decision errors. To test the inference capabilities of the proposed methods, in Fig. 4(a), we report the triangle error probability P_e , defined as the percentage of incorrectly estimated triangles with respect to the number of cliques with three edges in the simplex, versus the signal-to-noise ratio (SNR), when the observation contains only harmonic flows plus noise. We considered a simplex composed of $N = 50$ nodes and with a percentage of filled triangles with respect to the number of second order cliques in the graph equal to 50%. We also set $M = 50$, $t^* = 105$ and averaged our numerical results over 10^3 zero-mean signal and noise random

realizations. The harmonic signal bandwidth $|\mathcal{F}_H|$ is chosen equal to 105, which is equal to the dimension of the kernel of \mathbf{L}_1 . From Fig. 4(a), we can notice, as expected, that in the noiseless case the error probability is zero, since observing only harmonic flows enables perfect recovery of the matrix \mathbf{B}_2 . In the presence of noise, the MTV algorithm suffers and in fact we observe a non negligible error probability at low SNR. However, applying the PCA-BFMTV algorithm enables a significant recovery of performance, as evidenced by the blue curve that is entirely superimposed to the red curve, at least for the SNR values shown in the figure. In this example, the covariance matrix was estimated over 10^5 independent observations of the edge signals. The optimal γ coefficient was chosen after a cross validation operation following a line search approach aimed to minimize the error probability. The improvement of the PCA-BFMTV method with respect to the MTV method is due to the denoising made possible by the projection of the observed signal onto the space spanned by the eigenvectors associated with the largest eigenvalues of the estimated covariance matrix.

To test the proposed methods in the case where the observed signal contains both the solenoidal and harmonic components, in Fig. 4(b) we report P_e versus the SNR, for different values of the dimension of the subspace associated with the solenoidal part, indicated as $|\mathcal{F}_{\text{sol}}|$. From Fig. 4(b), we can observe that the performance of both algorithms MTV and PCA-BFMTV suffers when the bandwidth $|\mathcal{F}_{\text{sol}}|$ of the solenoidal component is large, whereas the performance degradation becomes negligible when $|\mathcal{F}_{\text{sol}}|$ is small. In all cases, PCA-BFMTV significantly outperforms the MTV algorithm, especially at low SNR values, because of its superior noise attenuation capabilities. As further numerical test, we run algorithm \mathcal{P}_S^k replacing the quadratic regularization term with the triple-wise coupling regularization function in (30), by obtaining the same performance of the PCA-BFMTV algorithm.

Performance on real data: The real data set we used to test our algorithms is the set of mobile phone calls collected in the city of Milan, Italy, by Telecom Italia, in the context of the Telecom Big Data Challenge [66]. The data are associated with a regular two-dimensional grid, composed of 100×100 points, superimposed to the city. Every point in the grid represents a square, of size 235 meters. In particular, the data set collects the number N_{ij} of calls from node i to area j , as a function of time. There is an edge between nodes i and j only if there is a non null traffic between those points. The traffic has been aggregated in time, over time intervals of one hour. We define the flow signal over edge (i, j) as $\Phi_{ij} = N_{ij} - N_{ji}$. We map all the values of matrix Φ into a vector of flow signals \mathbf{x}^1 . We observed the calls daily traffic during the month of December 2013. The data are aggregated for each day over an interval of one hour. Our first objective is to show whether there is an advantage in associating to the observed data set \mathbf{X}^1 a complex of order 2, i.e. a set of triangles, or it is sufficient to use a purely graph-based approach. In both cases, we rely on the same graph structure, whose \mathbf{B}_1 comes from the data set, after an arbitrary choice of the edges' orientation. If we use a graph-based approach, we can build a basis of the observed flow signals using the eigenvectors of the so called edge Laplacian

in [35], i.e. $\mathbf{L}_1^{\text{low}} = \mathbf{B}_1^T \mathbf{B}_1$. We call this basis $\mathbf{U}_1^{\text{low}}$. As an alternative, our proposed approach is to build a basis using the eigenvectors of $\mathbf{L}_1 = \mathbf{B}_1^T \mathbf{B}_1 + \mathbf{B}_2 \mathbf{B}_2^T$, where \mathbf{B}_2 is estimated from the data set \mathbf{X}^1 using our MTV algorithm. We call this basis \mathbf{U}_1 . To test the relative benefits of using \mathbf{U}_1 as opposed to $\mathbf{U}_1^{\text{low}}$, we run a basis pursuit algorithm with the goal of finding a good trade-off between the sparsity of the representation and the fitting error. More specifically, for any given observed vector $\mathbf{x}^1(m)$, we look for the sparse vector \mathbf{s}^1 as solution of the following basis pursuit problem [67]:

$$\begin{aligned} \min_{\mathbf{s}^1 \in \mathbb{R}^E} \quad & \|\mathbf{s}^1\|_1 & (\mathcal{B}) \\ \text{s.t.} \quad & \|\mathbf{x}^1 - \mathbf{V}\mathbf{s}^1\|_F \leq \epsilon \end{aligned} \quad (64)$$

where $\mathbf{V} = \mathbf{U}_1$ in our case, while $\mathbf{V} = \mathbf{U}_1^{\text{low}}$ in the graph-based approach. As a numerical result, in Fig. 5 we report the sparsity of the recovered edge signals versus the mean estimation error $\|\mathbf{x}^1 - \mathbf{V}\mathbf{s}^1\|_F$ considering as signal dictionary \mathbf{V} the eigenvectors of either the first-order Laplacian or the lower Laplacian. We used the MTV algorithm to infer the upper Laplacian matrix by setting the number t^* of triangles that we may detect equal to 800. This value is derived numerically through cross-validation over a training data set, by choosing the value of t^* that yields the minimum norm $\|\mathbf{s}^1\|_1$. As can be observed from Fig. 5, using the set of the eigenvectors of \mathbf{L}_1 yields a much smaller MSE, for a given sparsity or, conversely, a much more sparse representation, for a given MSE. An intuitive reason why our method performs so much better than a purely graph-based approach is that the matrix \mathbf{L}_1 has a much reduced kernel space with respect to $\mathbf{L}_1^{\text{low}}$ and the basis built on \mathbf{L}_1 captures much better some inner structure present in the data by inferring the structure of the additional term \mathbf{B}_2 from the data itself.

As a further test, we tested the two basis \mathbf{U}_1 and $\mathbf{U}_1^{\text{low}}$ in terms of the capability to recover the entire flow signal from a subset of samples. To this end, we exploit the band-limited property enforced by the sparse representation, enabling the use of the theory developed in Section V.A. Starting with the representation of each input vector \mathbf{x}^1 as $\mathbf{x}^1 = \mathbf{V}\mathbf{s}^1$, with either $\mathbf{V} = \mathbf{U}_1$ in our case, or $\mathbf{V} = \mathbf{U}_1^{\text{low}}$ in the graph-based approach, we used the Max-Det greedy sampling strategy in [56] to select the subset of edges where to sample the flow signal and then we used the recovery rule in (49) to retrieve the overall flow signal from the samples. The numerical results are reported in Fig. 6, representing the normalized recovering error of the edge signal versus the number N_s of samples used to reconstruct the overall signal. We can notice how introducing the term $\mathbf{B}_2 \mathbf{B}_2^T$, we can achieve a much smaller error, for the same number of samples.

VIII. CONCLUSION

In this paper we have presented an algebraic framework to analyze signals residing over a simplicial complex. In particular, we focused on signals defined over the edges of a complex of order two, i.e. including triangles, and we showed that exploiting the full algebraic structure of this complex provides an advantage with respect to graph-based tools only. We have not analyzed higher order signals. Nevertheless,

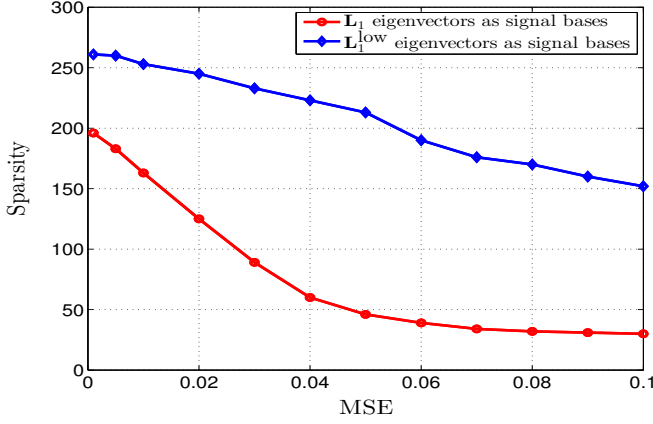


Fig. 5: Sparsity versus mean squared error.

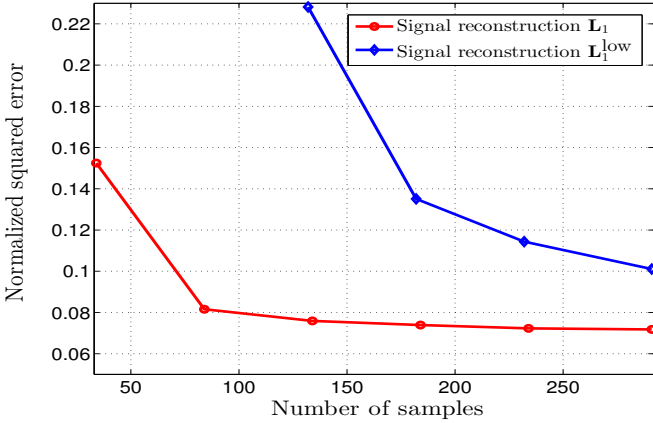


Fig. 6: Reconstruction error versus number of samples.

looking at the structure of the higher order Laplacian and to the Hodge decomposition, the tools derived in this paper can be directly translated to analyze signals defined over higher order structures. What would be missing in the higher order cases would be a visual interpretation of the properties of the higher order Laplacian eigenvectors, as we could not be talking about solenoidal or irrotational behaviors.

We proposed a method to infer the structure of a second order simplicial complex from flow data and we showed that, in applications over real wireless traffic data, the proposed approach can significantly outperform methods based only on graph representations. Furthermore, we proposed a method to analyze discrete vector fields and showed an application to the recovery of the RNA velocity field to predict the evolution of living cells. In such a case, using the eigenvectors of \mathbf{L}_1 we have been able to highlight irrotational and solenoidal behaviors that would have been difficult to highlight using only the eigenvectors of \mathbf{L}_0 . Further developments should include both theoretical aspects, especially in the statistical modeling of random variables defined over a simplicial complex, and the generalization to higher order structures.

APPENDIX A PROOF OF THEOREM 1

We begin by briefly recalling the basic properties of Lovász extension [49], [68] and then we proceed to the proof of Theorem 1.

Definition 1: Given a set \mathcal{A} , of cardinality $A = |\mathcal{A}|$, and its power set $2^{\mathcal{A}}$, let us consider a set function $Q : 2^{\mathcal{A}} \rightarrow \mathbb{R}$, with $Q(\emptyset) = 0$. Let $\mathbf{x} \in \mathbb{R}^A$ be a vector of real variables, ordered w.l.o.g. in increasing order, such that $x_1 \leq x_2 \leq \dots \leq x_A$. Define $C_0 \triangleq \mathcal{A}$ and $C_i \triangleq \{j \in \mathcal{A} : x_j > x_i\}$ for $i > 0$. Then, the Lovász extension $f : \mathbb{R}^A \rightarrow \mathbb{R}$ of Q , evaluated at \mathbf{x} , is given by [49]:

$$f(\mathbf{x}) = Q(\mathcal{A})x_1 + \sum_{i=1}^{A-1} Q(C_i)(x_{i+1} - x_i). \quad (65)$$

An interesting property of set functions is submodularity, defined as follows:

Definition 2: A set function $Q : 2^{\mathcal{A}} \rightarrow \mathbb{R}$ is *submodular* if and only if, $\forall \mathcal{B}, \mathcal{C} \subseteq \mathcal{A}$, it satisfies the following inequality:

$$Q(\mathcal{B}) + Q(\mathcal{C}) \geq Q(\mathcal{B} \cup \mathcal{C}) + Q(\mathcal{B} \cap \mathcal{C}).$$

A fundamental property is that a set function Q is submodular iff its Lovász extension $f(\mathbf{x})$ is a convex function and that minimizing $f(\mathbf{x})$ on $[0, 1]^A$, is equivalent to minimizing the set function Q on $2^{\mathcal{A}}$ [49, p.172]. Now, we wish to apply Lovász extension to a set function, defined over the edge set \mathcal{E} , counting the number of triangles gluing nodes belonging to a tri-partition of the node set \mathcal{V} . To this end, let us consider a tri-partition $\{\mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2\}$ of the node set \mathcal{V} . Extending the approach used for graphs, where the cut-size is introduced as a set function evaluated on the node set, to define a triangle-cut size, we need to introduce a set function defined on the edge set. Given a simplex \mathcal{X} , assume the faces of \mathcal{X} to be positively oriented with the order of their vertices, i.e. for the k -simplex σ_i^k we have $\sigma_i^k = [\sigma_{i_0}^{k-1}, \sigma_{i_1}^{k-1}, \dots, \sigma_{i_k}^{k-1}]$ with $i_0 < i_1 < \dots < i_k$. Given any non-empty tri-partition $\{\mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2\}$ of the node set \mathcal{V} , we define

$$\tau(\mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2) = \{\sigma_i^2 \in \mathcal{T} : |\sigma_i^2 \cap \mathcal{A}_j| = 1, 0 \leq j \leq 2\} \quad (66)$$

as the set of triangles $\sigma_i^2 \in \mathcal{T}$ with exactly one vertex in each set \mathcal{A}_j , for $j = 0, 1, 2$. Given any $\sigma_i^2 \in \tau(\mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2)$, there exists a permutation π_i of its vertices $\sigma_{i_j}^0$, $j = 0, 1, 2$, such that $\sigma_i^2 = [\sigma_{i_0}^0, \sigma_{i_1}^0, \sigma_{i_2}^0]$ and $\sigma_{i_j}^0 \in \mathcal{A}_j$, for $j = 0, 1, 2$. This returns a linear ordering¹ of the vertices of the partition such that for all $i < j$, if $\sigma_k^0 \in \mathcal{A}_i$ and $\sigma_m^0 \in \mathcal{A}_j$, then σ_k^0 precedes σ_m^0 in the vertex ordering. Given the tri-partition $\{\mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2\}$, we define the set \mathcal{E}_T of oriented edges such that an oriented edge $[\sigma_i^0, \sigma_j^0] \in \mathcal{E}_T$ if and only if $\sigma_i^0 \in \mathcal{A}_n$ and $\sigma_j^0 \in \mathcal{A}_m$ for some $n < m$. Thus, we can think of the triangle cut size F_1 in (29), equivalently, as a different function $G(\mathcal{E}_T, \overline{\mathcal{E}}_T)$ defined on sets of oriented edges, where $\overline{\mathcal{E}}_T$ denotes the complement set of \mathcal{E}_T in \mathcal{E} . Given an orientation on the simplicial complex, we can associate with the set of oriented edges \mathcal{E}_T the vector $\mathbf{1}_{\mathcal{E}_T} \in \mathbb{R}^E$, whose entries are 0 for edges in $\overline{\mathcal{E}}_T$ and ± 1 for edges in

¹A binary relation \prec on a set X is a *linear ordering* on X if it is transitive, antisymmetric and for any two elements $x, y \in X$, either $x \prec y$ or $y \prec x$.

\mathcal{E}_T , depending on the edge orientation. It is straightforward to check that $\mathbf{B}_2^T \mathbf{1}_{\mathcal{E}_T}$, where \mathbf{B}_2 is the edge-triangle incidence matrix defined in (3), computes the triangle cut for \mathcal{E}_T in the sense that each entry of this vector, corresponding to a triangle, is nonzero if and only if every vertex of that triangle is in a different set of the partition. As an example, considering the simplicial complex in Fig. 1, we have $\mathcal{E}_T = \{e_2, e_3, e_4, e_7, e_8\}$ and the associated vector $\mathbf{1}_{\mathcal{E}_T}$ is $[0, 1, 1, 1, 0, 0, 1, 1, 0, 0, 0]^T$. Thus, using (21), we get $\mathbf{B}_2^T \mathbf{1}_{\mathcal{E}_T} = [0, 1, 0]^T$ and the triangle cut size is equal to 1. Then, introducing the vector $\mathbf{t}_{\mathcal{E}_T} := \mathbf{B}_2^T \mathbf{1}_{\mathcal{E}_T}$, we define the edge set function as

$$G(\mathcal{E}_T, \bar{\mathcal{E}}_T) = \sum_{j=1}^T |\mathbf{t}_{\mathcal{E}_T}(j)|. \quad (67)$$

It can be proved that G is invariant under any permutation of the vertices of the triangles, since the effect of any vertex permutation is a change of sign in the entries of $\mathbf{t}_{\mathcal{E}_T}$. Therefore, we get

$$G(\mathcal{E}_T, \bar{\mathcal{E}}_T) = |\tau(\mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2)| = F_1(\mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2). \quad (68)$$

Now we wish to derive the Lovász extension of the set function $G(\mathcal{E}_T, \bar{\mathcal{E}}_T)$. Denote with $\sigma_i^2 = [\sigma_{i_0}^1, \sigma_{i_1}^1, \sigma_{i_2}^1]$ the oriented 2-order simplex where $1 \leq i_k \leq E$, for $k = 0, 1, 2$. To make explicit the dependence of the function $G(\mathcal{E}_T, \bar{\mathcal{E}}_T)$ on the edge indexes i_0, i_1, i_2 , we rewrite (67) as

$$G(\mathcal{E}_T, \bar{\mathcal{E}}_T) = \sum_{\sigma_i^2 \in \mathcal{S}} \left| \tilde{G}_i(\mathcal{E}_T, \bar{\mathcal{E}}_T) \right| \quad (69)$$

where, exploiting the structure of \mathbf{B}_2 , given in (3), each term \tilde{G}_i can be written as (we omit the dependence of \tilde{G}_i on the edge set partition, for notation simplicity)

$$\tilde{G}_i = \mathbf{1}_{\mathcal{E}_T}(\sigma_{i_0}^1) - \mathbf{1}_{\mathcal{E}_T}(\sigma_{i_1}^1) + \mathbf{1}_{\mathcal{E}_T}(\sigma_{i_2}^1), \quad (70)$$

where $\mathbf{1}_{\mathcal{E}_T}(\sigma_{i_k}^1)$, $k = 0, 1, 2$, are the entries of $\mathbf{1}_{\mathcal{E}_T}$ corresponding to the edge $\sigma_{i_k}^1$. \tilde{G}_i is then a set function defined only on the power set of $\{i_0, i_1, i_2\}$. We can now derive its Lovász extension $f(x_{i_0}, x_{i_1}, x_{i_2})$. First, assume $x_{i_0} \leq x_{i_1} \leq x_{i_2}$. Then, from Def. 1, we have $C_0 = \{i_0, i_1, i_2\}$, $C_1 = \{i_1, i_2\}$ and $C_2 = \{i_2\}$. Therefore, from (70), it holds:

$$\begin{aligned} \tilde{G}_i(C_0) &= \mathbf{1}_{\mathcal{E}_T}(\sigma_{i_0}^1) - \mathbf{1}_{\mathcal{E}_T}(\sigma_{i_1}^1) + \mathbf{1}_{\mathcal{E}_T}(\sigma_{i_2}^1) = 1 \\ \tilde{G}_i(C_1) &= -\mathbf{1}_{\mathcal{E}_T}(\sigma_{i_1}^1) + \mathbf{1}_{\mathcal{E}_T}(\sigma_{i_2}^1) = 0 \\ \tilde{G}_i(C_2) &= \mathbf{1}_{\mathcal{E}_T}(\sigma_{i_2}^1) = 1. \end{aligned} \quad (71)$$

Then, from (65) we get:

$$\begin{aligned} f(x_{i_0}, x_{i_1}, x_{i_2}) &= \tilde{G}_i(C_0)x_{i_0} + \tilde{G}_i(C_1)(x_{i_1} - x_{i_0}) \\ &\quad + \tilde{G}_i(C_2)(x_{i_2} - x_{i_1}) = x_{i_0} - x_{i_1} + x_{i_2}. \end{aligned}$$

Let us now assume $x_{i_1} \leq x_{i_0} \leq x_{i_2}$. Thereby, we have $C_0 = \{i_0, i_1, i_2\}$, $C_1 = \{i_0, i_2\}$ and $C_2 = \{i_2\}$, so that it results:

$$\begin{aligned} \tilde{G}_i(C_0) &= \mathbf{1}_{\mathcal{E}_T}(\sigma_{i_0}^1) - \mathbf{1}_{\mathcal{E}_T}(\sigma_{i_1}^1) + \mathbf{1}_{\mathcal{E}_T}(\sigma_{i_2}^1) = 1 \\ \tilde{G}_i(C_1) &= \mathbf{1}_{\mathcal{E}_T}(\sigma_{i_0}^1) + \mathbf{1}_{\mathcal{E}_T}(\sigma_{i_2}^1) = 2 \\ \tilde{G}_i(C_2) &= \mathbf{1}_{\mathcal{E}_T}(\sigma_{i_2}^1) = 1 \end{aligned} \quad (72)$$

and

$$f(x_{i_0}, x_{i_1}, x_{i_2}) = 2(x_{i_0} - x_{i_1}) + (x_{i_2} - x_{i_0}) + x_{i_1} = x_{i_0} - x_{i_1} + x_{i_2}.$$

By following similar derivations, it is not difficult to show that for $x_{i_0} \leq x_{i_2} \leq x_{i_1}$, $x_{i_1} \leq x_{i_2} \leq x_{i_0}$, $x_{i_2} \leq x_{i_1} \leq x_{i_0}$ and $x_{i_2} \leq x_{i_0} \leq x_{i_1}$, it holds $f(x_{i_0}, x_{i_1}, x_{i_2}) = x_{i_0} - x_{i_1} + x_{i_2}$. Therefore, from (69), defining the edge signal $\mathbf{x}^1 = [x_1, \dots, x_E]^T$, we can write the Lovász extension of $G(\mathcal{E}_T, \bar{\mathcal{E}}_T)$ exploiting the additive property [49] as

$$f_1(\mathbf{x}^1) = \sum_{\sigma_i^2 \in \mathcal{S}} |x_{i_0}^1 - x_{i_1}^1 + x_{i_2}^1| \quad (73)$$

or, equivalently, as

$$f_1(\mathbf{x}^1) = \sum_{j=1}^T \left| \sum_{i=1}^E B_2(i, j) x_i^1 \right|. \quad (74)$$

From the equality $G(\mathcal{E}_T, \bar{\mathcal{E}}_T) = F_1(\mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2)$ in (68), we can state that $f_1(\mathbf{x}^1)$ is the Lovász extension of F_1 . This completes the proof of Theorem 1.

REFERENCES

- [1] S. Barbarossa, S. Sardellitti, and E. Ceci, "Learning from signals defined over simplicial complexes," in *2018 IEEE Data Sci. Work. (DSW)*, 2018, pp. 51–55.
- [2] 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.
- [3] A. Ortega, P. Frossard, J. Kovačević, J. M. F. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges, and applications," *Proc. of the IEEE*, vol. 106, no. 5, pp. 808–828, May 2018.
- [4] S. Klamt, U.-U. Haus, and F. Theis, "Hypergraphs and cellular networks," *PLoS Comput. Biol.*, vol. 5, no. 5, 2009.
- [5] O. T. Courtney and G. Bianconi, "Generalized network structures: The configuration model and the canonical ensemble of simplicial complexes," *Phys. Rev. E*, vol. 93, no. 6, pp. 062311, 2016.
- [6] C. Giusti, R. Ghrist, and D. S. Bassett, "Two's company, three (or more) is a simplex," *J. Comput. Neurosci.*, vol. 41, no. 1, pp. 1–14, 2016.
- [7] T. Shen, Z. Zhang, Z. Chen, D. Gu, S. Liang, Y. Xu, R. Li, Y. Wei, Z. Liu, Y. Yi, and X. Xie, "A genome-scale metabolic network alignment method within a hypergraph-based framework using a rotational tensor-vector product," *Sci. Rep.*, vol. 8, no. 1, pp. 1–16, 2018.
- [8] A. R. Benson, R. Abebe, M. T. Schaub, A. Jadbabaie, and J. Kleinberg, "Simplicial closure and higher-order link prediction," *Proc. of Nat. Acad. of Sci.*, vol. 115, no. 48, pp. E11221–E11230, 2018.
- [9] S. Agarwal, K. Branson, and S. Belongie, "Higher order learning with graphs," in *Proc. 23rd Int. Conf. on Mach. Learn.*, 2006, pp. 17–24.
- [10] J. R. Munkres, *Topology*, Prentice Hall, 2000.
- [11] L. Kanari, P. Dłotko, M. Scolamiero, R. Levi, J. Shillcock, K. Hess, and H. Markram, "A topological representation of branching neuronal morphologies," *Neuroinform.*, vol. 16, no. 1, pp. 3–13, 2018.
- [12] A. Patania, G. Petri, and F. Vaccarino, "The shape of collaborations," *EPJ Data Sci.*, vol. 6, no. 1, pp. 18, 2017.
- [13] R. Ramanathan, A. Bar-Noy, P. Basu, M. Johnson, W. Ren, A. Swami, and Q. Zhao, "Beyond graphs: Capturing groups in networks," in *Proc. IEEE Conf. Comp. Commun. Work. (INFOCOM)*, 2011, pp. 870–875.
- [14] T. J. Moore, R. J. Drost, P. Basu, R. Ramanathan, and A. Swami, "Analyzing collaboration networks using simplicial complexes: A case study," in *Proc. IEEE Conf. Comp. Commun. Work. (INFOCOM)*, 2012, pp. 238–243.
- [15] T. Roman, A. Nayyeri, B. T. Fasy, and R. Schwartz, "A simplicial complex-based approach to unmixing tumor progression data," *BMC Bioinform.*, vol. 16, no. 1, pp. 254, 2015.
- [16] G. Carlsson, "Topology and data," *Bulletin Amer. Math. Soc.*, vol. 46, no. 2, pp. 255–308, 2009.
- [17] A. Muhammad and M. Egerstedt, "Control using higher order Laplacians in network topologies," in *Proc. of 17th Int. Symp. Math. Theory Netw. Syst.*, 2006, pp. 1024–1038.
- [18] 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.

- [19] Q. Xu, Q. Huang, T. Jiang, B. Yan, W. Lin, and Y. Yao, "HodgeRank on random graphs for subjective video quality assessment," *IEEE Trans. Multimedia*, vol. 14, no. 3, pp. 844–857, June 2012.
- [20] A. Tahbaz-Salehi and A. Jadbabaie, "Distributed coverage verification in sensor networks without location information," *IEEE Trans. Autom. Contr.*, vol. 55, no. 8, pp. 1837–1849, Aug. 2010.
- [21] H. Chintakunta and H. Krim, "Distributed localization of coverage holes using topological persistence," *IEEE Trans. Signal Process.*, vol. 62, no. 10, pp. 2531–2541, May 2014.
- [22] V. De Silva and R. Ghrist, "Coverage in sensor networks via persistent homology," *Algeb. & Geom. Topol.*, vol. 7, no. 1, pp. 339–358, 2007.
- [23] S. Emrani, T. Gentimis, and H. Krim, "Persistent homology of delay embeddings and its application to wheeze detection," *IEEE Signal Process. Lett.*, vol. 21, no. 4, pp. 459–463, Apr. 2014.
- [24] H. Edelsbrunner and J. Harer, "Persistent homology—a survey," *Contemp. Math.*, vol. 453, pp. 257–282, 2008.
- [25] D. Horak, S. Maletić, and M. Rajković, "Persistent homology of complex networks," *J. Statist. Mech.: Theory and Experiment*, vol. 2009, no. 3, pp. P03034, 2009.
- [26] H. Krim and A. B. Hamza, *Geometric methods in signal and image analysis*, Cambridge Univ. Press, 2015.
- [27] M. Robinson, *Topological signal processing*, Springer, 2016.
- [28] W. Huang, T. A. W. Bolton, J. D. Medaglia, D. S. Bassett, A. Ribeiro, and D. Van De Ville, "A graph signal processing perspective on functional brain imaging," *Proc. of the IEEE*, vol. 106, no. 5, pp. 868–885, May 2018.
- [29] 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.
- [30] R. Sever and J. S. Brugge, "Signal transduction in cancer," *Cold Spring Harbor Perspec. in Medic.*, vol. 5, no. 4: a006098, 2015.
- [31] M. T. Schaub and S. Segarra, "Flow smoothing and denoising: Graph signal processing in the edge-space," in *2018 IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Nov. 2018, pp. 735–739.
- [32] J. Jia, M. T. Schaub, S. Segarra, and A. R. Benson, "Graph-based semi-supervised & active learning for edge flows," in *Proc. of 25th ACM SIGKDD Int. Conf. Knowl. Disc. & Data Min.*, 2019, pp. 761–771.
- [33] T. S. Evans and R. Lambiotte, "Line graphs, link partitions, and overlapping communities," *Phys. Rev. E*, vol. 80, Jul. 2009.
- [34] Y.-Y. Ahn, J. P. Bagrow, and S. Lehmann, "Link communities reveal multiscale complexity in networks," *Nat.*, vol. 466, pp. 761–764, 2010.
- [35] M. Mesbahi and M. Egerstedt, *Graph theoretic methods in multiagent networks*, vol. 33, Princ. Univ. Press, 2010.
- [36] S. Mukherjee and J. Steenbergen, "Random walks on simplicial complexes and harmonics," *Random Struct. & Algor.*, vol. 49, no. 2, pp. 379–405, 2016.
- [37] O. Parzanchevski and R. Rosenthal, "Simplicial complexes: spectrum, homology and random walks," *Random Struct. & Algor.*, vol. 50, no. 2, pp. 225–261, 2017.
- [38] M. T. Schaub, A. R. Benson, P. Horn, G. Lippner, and A. Jadbabaie, "Random walks on simplicial complexes and the normalized Hodge Laplacian," *arXiv preprint arXiv:1807.05044*, 2018.
- [39] 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.
- [40] X. Dong, D. Thanou, M. Rabbat, and P. Frossard, "Learning graphs from data: A signal representation perspective," *IEEE Signal Process. Mag.*, vol. 36, no. 3, pp. 44–63, May 2019.
- [41] G. La Manno et al., "RNA velocity of single cells," *Nature*, vol. 560, no. 7719, pp. 494–498, 2018.
- [42] L. J. Grady and J. R. Polimeni, *Discrete calculus: Applied analysis on graphs for computational science*, Sprin. Sci. & Busin. Media, 2010.
- [43] T. E. Goldberg, "Combinatorial Laplacians of simplicial complexes," *Senior Thesis, Bard College*, 2002.
- [44] U. Von Luxburg, "A tutorial on spectral clustering," *Stat. Comput.*, vol. 17, no. 4, pp. 395–416, 2007.
- [45] D. Horak and J. Jost, "Spectra of combinatorial Laplace operators on simplicial complexes," *Adv. Math.*, vol. 244, pp. 303–336, 2013.
- [46] J. Steenbergen, *Towards a spectral theory for simplicial complexes*, Ph.D. thesis, Duke Univ., 2013.
- [47] L.-H. Lim, "Hodge Laplacians on graphs," *S. Mukherjee (Ed.), Geometry and Topology in Statistical Inference, Proc. Sympos. Appl. Math.*, 76, AMS, 2015.
- [48] B. Eckmann, "Harmonische funktionen und randwertaufgaben in einem komplex," *Comment. Math. Helv.*, vol. 17, no. 1, pp. 240–255, 1944.
- [49] F. Bach, *Learning with Submodular Functions: A Convex Optimization Perspective*, vol. 6, Foundations and Trends in Machine Learning, 2013.
- [50] L. Lovász, "Submodular functions and convexity," in *A. Bachem et al. (eds.) Math. Program. The State of the Art, Springer Berlin Heidelberg*, pp. 235–257, 1983.
- [51] S. Sardellitti, S. Barbarossa, and P. Di Lorenzo, "On the graph Fourier transform for directed graphs," *IEEE J. Sel. Top. Signal Process.*, vol. 11, no. 6, pp. 796–811, Sept. 2017.
- [52] O. Parzanchevski, R. Rosenthal, and R. J. Tessler, "Isoperimetric inequalities in simplicial complexes," *Combinatorica*, vol. 36, no. 2, pp. 195–227, 2016.
- [53] J. Steenbergen, C. Klivans, and S. Mukherjee, "A Cheeger-type inequality on simplicial complexes," *Adv. Appl. Math.*, vol. 56, pp. 56–77, 2014.
- [54] M. Hein, S. Setzer, L. Jost, and S. S. Rangapuram, "The total variation on hypergraphs - Learning on hypergraphs revisited," in *Adv. Neur. Inform. Proces. Syst. (NIPS)*, Dec. 5-8 2013.
- [55] S. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge Univ. Press, 2004.
- [56] M. Tsitsvero, S. Barbarossa, and P. Di Lorenzo, "Signals on graphs: Uncertainty principle and sampling," *IEEE Trans. Signal Process.*, vol. 64, no. 18, pp. 4845–4860, Sept. 2016.
- [57] S. Barbarossa and M. Tsitsvero, "An introduction to hypergraph signal processing," in *Proc. of IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2016, pp. 6425–6429.
- [58] S. Barbarossa and S. Sardellitti, "Topological signal processing over simplicial complexes," *arXiv preprint arxiv.org/abs/1907.11577v2*, 2020.
- [59] Élie Cartan, "Sur certaines expressions différentielles et le problème de pfaff," in *Annales scientifiques de l'École Normale Supérieure*, 1899, vol. 16, pp. 239–332.
- [60] A. N. Hirani, *Discrete Exterior Calculus*, PhD thesis, California Institute of Technology, May 2003.
- [61] N. Bell and A. N. Hirani, "PyDEC: software and algorithms for discretization of exterior calculus," *ACM Trans. Math. Soft. (TOMS)*, vol. 39, no. 1, pp. 3, 2012.
- [62] M. Fisher, P. Schröder, M. Desbrun, and H. Hoppe, "Design of tangent vector fields," in *ACM Trans. Grap. (TOG)*, 2007, vol. 26, p. 56.
- [63] C. Brandt, L. Scandolo, E. Eiseemann, and K. Hildebrandt, "Spectral processing of tangential vector fields," in *Computer Graphics Forum*. Wiley Online Library, 2017, vol. 36, pp. 338–353.
- [64] J. Dodziuk, "Finite-difference approach to the Hodge theory of harmonic forms," *Amer. Jour. Math.*, vol. 98, no. 1, pp. 79–104, 1976.
- [65] G. B. Giannakis, Y. Shen, and G. V. Karanikolas, "Topology identification and learning over graphs: Accounting for nonlinearities and dynamics," *Proc. of the IEEE*, vol. 106, no. 5, pp. 787–807, May 2018.
- [66] *Big data-Dandelion API datasets*, available at: <https://dandelion.eu/>.
- [67] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," in *SIAM J. Sci. Comput.*, 1998, vol. 20, pp. 33–61.
- [68] L. Jost, S. Setzer, and M. Hein, "Nonlinear eigenproblems in data analysis: Balanced graph cuts and the ratioDCA-Prox," in *Extrac. Quant. Infor. Complex Syst., Spr. Intern. Publ.*, vol. 102, pp. 263–279, 2014.



Sergio Barbarossa (S'84-M'88-F'12) is a Full Professor and Senior Research Fellow at the Sapienza School of Advanced Studies, Sapienza University of Rome, Italy. He is an IEEE Fellow and EURASIP Fellow. He received the 2000 and 2014 IEEE Best Paper Awards from the IEEE Signal Processing Society and the 2010 Technical Achievements Award from the EURASIP. His current research interests are in the fields of topological signal processing and edge machine learning.



Stefania Sardellitti (M'12) received the M.Sc. degree in Electronic Engineering from the 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 at the University of Cassino, Italy. Currently, she is assistant professor with the Department of Information Engineering, Electronics and Telecommunications, University of Rome, Sapienza, Italy. She received the 2014 IEEE Best

Paper Award from the IEEE Signal Processing Society. Her current research interests are in the area of graph signal processing, mobile edge computing and distributed optimization.