TOPOLOGICAL SLEPIANS: MAXIMALLY LOCALIZED REPRESENTATIONS OF SIGNALS OVER SIMPLICIAL COMPLEXES

Claudio Battiloro, Paolo Di Lorenzo, and Sergio Barbarossa

DIET Department, Sapienza University of Rome, Via Eudossiana 18, 00184, Rome, Italy
E-mail: {claudio.battiloro, paolo.dilorenzo, sergio.barbarossa}@uniroma1.it

ABSTRACT

This paper introduces topological Slepians, i.e., a novel class of signals defined over topological spaces (e.g., simplicial complexes) that are maximally concentrated on the topological domain (e.g., over a set of nodes, edges, triangles, etc.) and perfectly localized on the dual domain (e.g., a set of frequencies). These signals are obtained as the principal eigenvectors of a matrix built from proper localization operators acting over topology and frequency domains. Then, we suggest a principled procedure to build dictionaries of topological Slepians, which theoretically provide non-degenerate frames. Finally, we evaluate the effectiveness of the proposed topological Slepian dictionary in two applications, i.e., sparse signal representation and denoising of edge flows.

Index Terms— Topological signal processing, simplicial complexes, localized representation, sparse signal recovery, frames.

1. INTRODUCTION

In the last few years, there was a surge of interest in developing processing techniques for signals defined over irregular domains, which are not necessarily metric spaces. In particular, the field of graph signal processing (GSP) studies methodologies to analyze and process signals defined over the vertices of a graph [1], [2]. To this aim, several graph operators (e.g. adjacency, Laplacian, etc.) have been introduced, thus leading to alternative designs of filters on graphs and graph Fourier transforms. The main feature of these processing tools is that they come to depend on the connectivity of the graph, which is encoded into the structure of the adopted graph operator. However, despite their overwhelming popularity, graph representations can only take into account *pairwise* relationships among data.

In complex interconnected systems, the interactions often cannot be reduced to simple pairwise relationships, and graph representations might result incomplete and inefficient [3–9]. For instance, in biological networks, multi-way interactions among complex substances (such as genes, proteins, or metabolites) cannot be evoked using simply pairwise relationships [4]; also, in the brain, groups of neurons typically activate at the same time [6]. These applications have sparked a renewed interest in extending GSP tools to incorporate multi-way relationships among data, thus leading to the emergent field of topological signal processing (TSP) [10,11]. In this context, the seminal works [10, 11] illustrated the benefits obtained by processing signals defined over simplicial complexes, which are specific examples of hyper-graphs with a rich algebraic description that can easily encode multi-way data. Then, several papers have given important contributions to TSP. For instance, the work in [12] proposed FIR filters for signals defined over simplicial complexes, hinging on the Hodge decomposition, where the Fourier modes are eigenvectors of higher order combinatorial Laplacians [13]. The work in [14] introduced generalized Laplacian for embedding simplicial

complexes into traditional graphs. In [15] the authors proposed self-driven graph Volterra models that can capture higher-order interactions among nodal observables available in networked data. Very recently, some works focused on processing signals defined over cell complexes, i.e., topological spaces not constrained to respect any inclusion property [16, 17]. Finally, deep neural architectures able to learn from data defined over topological spaces (e.g., simplicial or cell complexes) have been recently developed in [18–23].

One of the fundamental problems in signal processing is to obtain a sparse representation of the signals of interest. Typically, the sparser is the representation, the better is the performance of several processing tasks such as, e.g., compression, denoising, sampling and recovery [24]. In the context of TSP, a natural basis for signal representation is given by the topological Fourier modes [13]. However, often the signal of interest is highly localized over a portion of the topological domain, i.e., it is mostly concentrated over a set of nodes, edges, or triangles, while being at the same time highly (if not perfectly) localized over a specific set of frequencies. In such a case, Fourier modes might not lead to an efficient and sparse representation of localized topological signals, since they are usually non-sparse, i.e., not localized on sub-regions of the complex. To find localized spectral representations, the work in [25] proposed a family of wavelets for simplicial signals, respecting the Hodge decomposition and extending the graph wavelets from [26].

In this work, moved by the necessity of finding efficient sparse representations of localized signals lying over simplicial complexes, we introduce topological Slepians, i.e., a class of signals that are maximally concentrated on the simplicial domain (e.g., over a set of nodes, edges, triangles, etc.) and perfectly localized on the dual domain (e.g., a set of frequencies). Due to their joint localization properties, they represent the simplicial counterpart of the prolate spheroidal wave functions introduced by Slepian and Pollack for continuous-time signals [27], and of the graph Slepians introduced in [28,29] for graph signals. These signals are obtained as the principal eigenvectors of a matrix built from localization operators defined over the topological space and its dual (frequency) domain. Then, we propose a principled way to build a dictionary having topological Slepians as atoms, identifying the theoretical conditions necessary to build a non-degenerate frame. Finally, we apply the proposed methodology to two cases of interest, namely, sparse signal representation and denoising of edge flows, illustrating how topological Slepians compare favourably with state-of-the-art techniques.

2. BACKGROUND

In this section, we review basics of topological signal processing over simplicial complexes that will be useful along the paper.

Simplicial complex and signals. Given a finite set of vertices $\mathcal{V}$, a k-simplex $\mathcal{H}_{k,i}$ is a subset of $\mathcal{V}$ with cardinality $k + 1$, e.g., edges are 1-simplices, and triangles are 2-simplices. A face of $\mathcal{H}_{k,i}$ is a subset with cardinality k, and thus a k-simplex has $k + 1$ faces. A

coface of $\mathcal{H}_{k,i}$ is a $(k+1)$-simplex that includes $\mathcal{H}_{k,i}$ [10, 30]. If two simplices share a common face, then they are lower neighbours; if they share a common coface, they are upper neighbours [31]. A simplicial complex $\mathcal{X}_K$ of order K, is a collection of k-simplices $\mathcal{H}_{k,i}, k = 0, \ldots, K$ such that, for any $\mathcal{H}_{k,i} \in \mathcal{X}_K, \mathcal{H}_{k-1,j} \in \mathcal{X}_K$ if $\mathcal{H}_{k-1,j} \subset \mathcal{H}_{k,i}$ (inclusivity property). We denote the set of k-simplices in $\mathcal{X}_K$ as $\mathcal{D}_k := \{\mathcal{H}_{k,i} : \mathcal{H}_{k,i} \in \mathcal{X}_K, i = 1, ..., N_k\}$, with $|\mathcal{D}_k| = N_k$ and $\mathcal{D}_k \subseteq \mathcal{X}_K$.

We are interested in processing signals defined over a simplicial complex. A k-simplicial signal is defined as a mapping from the set of all k-simplices contained in the complex to real numbers:

$$\mathbf{x}_k : \mathcal{D}_k \to \mathbb{R}, \quad k = 0, 1, \ldots K. \quad (1)$$

The order of the signal is one less the cardinality of the elements of $\mathcal{D}_k$. In most of the cases the focus is on complex $\mathcal{X}_2$ of order up to two, thus a set of vertices $\mathcal{V}$ with $|\mathcal{V}| = V$, a set of edges $\mathcal{E}$ with $|\mathcal{E}| = E$ and a set of triangles $\mathcal{T}$ with $|\mathcal{T}| = T$ are considered, resulting in $\mathcal{D}_0 = \mathcal{V}$ (simplices of order 0), $\mathcal{D}_1 = \mathcal{E}$ (simplices of order 1) and $\mathcal{D}_2 = \mathcal{T}$ (simplices of order 2).

Algebraic representations. The structure of a simplicial complex $\mathcal{X}_K$ is fully described by the set of its incidence matrices $\mathbf{B}_k, k = 1, \ldots, K$, given a reference orientation. The entries of the incidence matrix $\mathbf{B}_k$ establish which k-simplices are incident to which $(k-1)$-simplices. Denoting the fact that two simplex have the same orientation with $\mathcal{H}_{k-1,i} \sim \mathcal{H}_{k,j}$ and viceversa with $\mathcal{H}_{k-1,i} \nsim \mathcal{H}_{k,j}$, the entries of $\mathbf{B}_k$ are defined as:

$$\left[\mathbf{B}_k\right]_{i,j} = \begin{cases} 0, & \text{if } \mathcal{H}_{k-1,i} \not\subset \mathcal{H}_{k,j} \\ 1, & \text{if } \mathcal{H}_{k-1,i} \subset \mathcal{H}_{k,j} \text{ and } \mathcal{H}_{k-1,i} \sim \mathcal{H}_{k,j} \\ -1, & \text{if } \mathcal{H}_{k-1,i} \subset \mathcal{H}_{k,j} \text{ and } \mathcal{H}_{k-1,i} \nsim \mathcal{H}_{k,j} \end{cases} \quad (2)$$

From the incidence information, we can build the high order combinatorial Laplacian matrices [32] of order $k = 0, \ldots, K$ as:

$$\mathbf{L}_0 = \mathbf{B}_1 \mathbf{B}_1^T, \quad (3)$$

$$\mathbf{L}_k = \underbrace{\mathbf{B}_k^T \mathbf{B}_k}_{\mathbf{L}_k^d} + \underbrace{\mathbf{B}_{k+1} \mathbf{B}_{k+1}^T}_{\mathbf{L}_k^u}, \quad k = 1, \ldots, K-1, \quad (4)$$

$$\mathbf{L}_K = \mathbf{B}_K^T \mathbf{B}_K. \quad (5)$$

All Laplacian matrices of intermediate order, i.e. $k = 1, \ldots, K-1$, contain two terms: The first term $\mathbf{L}_k^d$, also known as lower Laplacian, encodes the lower connectivity among k-order simplices; the second term $\mathbf{L}_k^u$, also known as upper Laplacian, encodes the upper connectivity among k-order simplices. 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.

Hodge decomposition: High order Laplacians admit a Hodge decomposition [30], such that the k-simplicial signal space can be decomposed as:

$$\mathbb{R}^{N_k} = \mathbf{im}\left(\mathbf{B}_k^T\right) \bigoplus \mathbf{im}\left(\mathbf{B}_{k+1}\right) \bigoplus \mathbf{ker}\left(\mathbf{L}_k\right), \quad (6)$$

where $\bigoplus$ is the direct sum of vector spaces, and $\mathbf{ker}(\cdot)$ and $\mathbf{im}(\cdot)$ are the kernel and image spaces of a matrix, respectively. Thus, any signal $\mathbf{x}_k$ of order k admits the following orthogonal decomposition:

$$\mathbf{x}_k = \underbrace{\mathbf{B}_k^T \mathbf{x}_{k-1}}_{(a)} + \underbrace{\mathbf{B}_{k+1} \mathbf{x}_{k+1}}_{(b)} + \underbrace{\widetilde{\mathbf{x}}_k}_{(c)}. \quad (7)$$

To give an interpretation of (7), let us consider the decomposition of edge flow signals $\mathbf{x}_1$, i.e., $k = 1$. Then, the following holds [10, 33]:

(a) Applying matrix $\mathbf{B}_1$ to an edge flow $\mathbf{x}_1$ computes its net flow (i.e. the difference between inflow and outflow) at each node, while applying its adjoint $\mathbf{B}_1^T$ to a node signal $\mathbf{x}_0$ computes the gradients along the edges. We call $\mathbf{B}_1^T \mathbf{x}_0$ the *irrotational component* of $\mathbf{x}_1$ and $\mathbf{im}(\mathbf{B}_1^T)$ the *gradient space*.

(b) The matrix $\mathbf{B}_2^T$ is a curl operator and applying it to an edge flow $\mathbf{x}_1$ computes the circulation over each triangle. Its adjoint $\mathbf{B}_2$ induces an edge flow $\mathbf{x}_1$ from a triangle signal $\mathbf{x}_2$. We call $\mathbf{B}_2 \mathbf{x}_2$ the *solenoidal component* of $\mathbf{x}_1$ and $\mathbf{im}(\mathbf{B}_2)$ the *curl space*.

(c) The remaining component $\widetilde{\mathbf{x}}_1$ is called the *harmonic component* and $\mathbf{ker}(\mathbf{L}_1)$ is called the harmonic space. Any edge flow $\widetilde{\mathbf{x}}_1$ has zero divergence and zero curl.

Simplicial Fourier transform. Simplicial signals of various order can be represented over the bases of the eigenvectors of the corresponding high order Laplacian matrices. Hence, using the eigendecomposition $\mathbf{L}_k = \mathbf{U}_k \mathbf{\Lambda}_k \mathbf{U}_k^T$, the Simplicial Fourier Transform (SFT) of order k is defined as the projection of a k-order signal onto the eigenvectors of $\mathbf{L}_k$ [10, 31]:

$$\widehat{\mathbf{x}}_k \triangleq \mathbf{U}_k^T \mathbf{x}_k. \quad (8)$$

We refer to the eigenvalue domain of the SFT as the frequency domain. An interesting consequence of the Hodge decomposition in (7) is that the eigenvectors belonging to $\mathbf{im}(\mathbf{L}_k^d)$ are orthogonal to those belonging to $\mathbf{im}(\mathbf{L}_k^u)$, for all $k = 1, \ldots, K-1$. Therefore, the eigenvectors of $\mathbf{L}_k$ are given by the union of the eigenvectors of $\mathbf{L}_k^u$ (spanning the curl sub-space), the eigenvectors of $\mathbf{L}_k^d$ (spanning the gradient sub-space), and the kernel of $\mathbf{L}_k$ (spanning the harmonic sub-space) [31]. For the sake of simplicity, but without loss of generality, in the sequel we will focus on the processing of edge flow signals. Thus, we will denote $\mathbf{x}_1$ with $\mathbf{x}$, $\mathbf{U}_k$ with $\mathbf{U}$, $\mathbf{L}_1$ with $\mathbf{L}$, $\mathbf{L}_1^d$ with $\mathbf{L}^d$ and $\mathbf{L}_1^u$ with $\mathbf{L}^u$, such that $\mathbf{L} = \mathbf{L}^d + \mathbf{L}^u$.

3. TOPOLOGICAL SLEPIANS

The eigenvectors of higher order Laplacians composing the simplicial Fourier basis are generally non-sparse, meaning that exploiting them for sparse representation of localized signals usually leads to poor performance [25]. Thus, in this section we introduce *topological Slepians*, i.e., a novel class of simplicial signals that are maximally concentrated over a sub-set of edges, while being perfectly localized over the dual domain, i.e. a set of frequencies. Following the approach of [28], we introduce two localization operators acting onto an edge concentration set, say $\mathcal{S}$, and onto a frequency concentration set, say $\mathcal{F}$, respectively. The edge-limiting operator onto the edge set $\mathcal{S}$ is defined as the matrix $\mathbf{C}_\mathcal{S} \in \mathbb{R}^{E \times E}$ given by:

$$\mathbf{C}_\mathcal{S} = \text{diag}(\mathbf{1}_\mathcal{S}), \quad (9)$$

where $\mathbf{1}_\mathcal{S} \in \mathbb{R}^E$ is a vector having ones in the index positions specified in $\mathcal{S}$, and zero otherwise; and $\text{diag}(\mathbf{z})$ denotes a diagonal matrix having $\mathbf{z}$ on the diagonal. Clearly, from (9), an edge signal $\mathbf{x}$ is perfectly localized onto the set $\mathcal{S}$ if $\mathbf{C}_\mathcal{S} \mathbf{x} = \mathbf{x}$. Similarly, the frequency limiting operator can be defined as:

$$\mathbf{B}_\mathcal{F} = \mathbf{U} \, \text{diag}(\mathbf{1}_\mathcal{F}) \, \mathbf{U}^T, \quad (10)$$

which represents an ideal band-pass filter over the frequency set $\mathcal{F}$. Clearly, an edge signal is perfectly localized over the bandwidth $\mathcal{F}$ if $\mathbf{B}_\mathcal{F} \mathbf{x} = \mathbf{x}$. The matrices in (9) and (10) are projection operators, having maximum spectral radius equal to one.

We define topological Slepians as the set of orthonormal vectors that are maximally concentrated over the edge set $\mathcal{S}$, and perfectly

localized onto the bandwidth $\mathcal{F}$. Mathematically, topological Slepians are given by the solution of the following optimization problem:

$$\boldsymbol{\psi}_i = \arg\max_{\boldsymbol{\psi}_i} \ ||\mathbf{C}_{\mathcal{S}}\boldsymbol{\psi}_i||_2^2$$

$$\text{subject to } ||\boldsymbol{\psi}_i|| = 1, \quad \mathbf{B}_{\mathcal{F}}\boldsymbol{\psi}_i = \boldsymbol{\psi}_i, \qquad (11)$$
$$< \boldsymbol{\psi}_i, \boldsymbol{\psi}_j >= 0, \quad j = 1,\ldots,i-1, \text{ if } i > 1,$$

for all $i = 1,\ldots,E$. The solution of problem (11) is given by the following theorem, whose proof follows the same steps as in [28] and is omitted due to lack of space.

Theorem 1. The set of topological Slepians $\{\boldsymbol{\psi}_i\}_{i=1}^C$ with $C :=$ rank$\{\mathbf{B}_{\mathcal{F}}\mathbf{C}_{\mathcal{S}}\mathbf{B}_{\mathcal{F}}\}$ that solve problem (11) is given by the eigenvectors of the matrix operator $\mathbf{B}_{\mathcal{F}}\mathbf{C}_{\mathcal{F}}\mathbf{B}_{\mathcal{F}}$, i.e.,

$$\mathbf{B}_{\mathcal{F}}\mathbf{C}_{\mathcal{S}}\mathbf{B}_{\mathcal{F}}\boldsymbol{\psi}_i = \lambda_i\boldsymbol{\psi}_i, \qquad (12)$$

where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_C > 0$. Furthermore, it holds $< \boldsymbol{\psi}_i, \mathbf{C}_{\mathcal{S}}\boldsymbol{\psi}_j >= \lambda_j\delta_{i,j}$, where $\delta_{i,j}$ is the Kronecker delta. ∎

In Fig. 1, we illustrate the first three topological Slepians maximally concentrated over the edge set in Fig. 1 (a), and bandlimited over the gradient sub-space (i.e., the set of frequency indexes associated with the eigenvectors of $\mathbf{L}^d$). Here, we consider topological Slepians for edge flows (i.e., $k = 1$), but the approach can be similarly applied at any order $k = 0,\ldots,K$, properly defining the operators (9) and (10) over higher-order dual domains. In particular, the graph Slepians introduced in [28] represent a very special case of our general formulation, obtained for $k = 0$. Finally, the dual formulation of (11), aimed at finding maximally localized signals over a set of frequencies, while being perfectly localized over an edge set, can be readily obtained exchanging the roles of $\mathbf{C}_{\mathcal{S}}$ and $\mathbf{B}_{\mathcal{F}}$ in (11).

4. DICTIONARY OF TOPOLOGICAL SLEPIANS

In this section we present a principled way to build overcomplete dictionaries of topological Slepians, relying on high order Laplacians in (4) and the Hodge decomposition in (7). Let $\boldsymbol{\Psi}_{\mathcal{S},\mathcal{F}} = \{\boldsymbol{\psi}_i\}_{i=1}^C$ be the set of topological Slepians obtained from (12) considering the edge set $\mathcal{S}$ and the frequency set $\mathcal{F}$. In very general terms, a dictionary of topological Slepians has the following form:

$$\mathbf{D}_{\mathcal{C}} = \Big[\boldsymbol{\Psi}_{\mathcal{S}_1,\mathcal{F}_1}, ..., \boldsymbol{\Psi}_{\mathcal{S}_i,\mathcal{F}_i}, ..., \boldsymbol{\Psi}_{\mathcal{S}_M,\mathcal{F}_M}\Big], \qquad (13)$$

which collects M sets of topological Slepians obtained from the pairs of concentration sets $\{\mathcal{S}_i, \mathcal{F}_i\}_{i=1}^M$. Therefore, building a dictionary of topological Slepians translates into properly choosing a sequence of concentration sets $\mathcal{C} = \{\mathcal{S}_i, \mathcal{F}_i\}_{i=1}^M$, i.e., a proper way of covering the edge set and the frequency domain. Let us initially assume that the kernel of the Laplacian $\mathbf{L}$ is empty, i.e., there are no harmonic components in (7). Then, from (6), the frequency indexes can be divided into two separate sets: (i) the *irrotational band*, which corresponds to the set of eigenvalues of $\mathbf{L}^d$, say $\mathcal{F}^d$; and (ii) the *solenoidal band*, which corresponds to the set of eigenvalues of $\mathbf{L}^u$, say $\mathcal{F}^u$. Thus, a very natural way to partition the frequency indexes is to choose $\mathcal{F}^d$ and $\mathcal{F}^u$ as frequency concentration sets, leading to irrotational and solenoidal Slepian components, respectively. Regarding the edge concentration sets, the Hodge Decomposition in (7) and the Laplacian structure in (4) suggest us to define two distinct sequences of (not necessarily disjoint) edge concentration sets: (i) K^d sets $\{\mathcal{S}_i^d\}_{i=1}^{K^d}$ based on lower neighborhoods (encoded by $\mathbf{L}^d$); (ii) K^u sets $\{\mathcal{S}_i^u\}_{i=1}^{K^u}$ based on upper neighborhoods (encoded by $\mathbf{L}^u$). Since the solenoidal and irrotational bandwidths are related to the eigenvectors of $\mathbf{L}^u$ and $\mathbf{L}^d$ [cf. (7)], we propose to associate
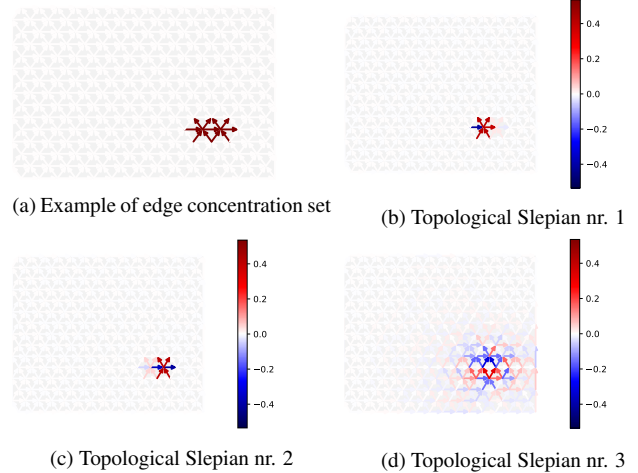

(a) Example of edge concentration set

(b) Topological Slepian nr. 1

(c) Topological Slepian nr. 2

(d) Topological Slepian nr. 3

Fig. 1: Example of topological Slepians on edge flows.

$\mathcal{F}^u$ to each upper concentration set $\mathcal{S}_i^u$, and $\mathcal{F}^d$ to each lower concentration set $\mathcal{S}_i^d$, such that the sequence $\mathcal{C}$ of concentration sets for the dictionary of topolgical Slepians in (13) reads as (with a slight abuse of notation):

$$\mathcal{C} = \{\mathcal{S}_i^u, \mathcal{F}^u\}_{i=1}^{K^u} \cup \{\mathcal{S}_i^d, \mathcal{F}^d\}_{i=1}^{K^d}. \qquad (14)$$

Finally, if the kernel of the Laplacian $\mathbf{L}$ is not empty, we propose to add to the dictionary $\mathbf{D}_{\mathcal{C}}$ also the eigenvectors $\overline{\mathbf{U}} = \{\overline{\mathbf{u}}_k\}_k$ of $\mathbf{L}$ associated with the zero eigenvalues. Therefore, given a sequence $\mathcal{C}$ of concentration sets built as in (14), the corresponding dictionary of topological Slepians reads as:

$$\mathbf{D}_{\mathcal{C}} = \Big[\boldsymbol{\Psi}_{\mathcal{S}_1^u,\mathcal{F}^u}, ..., \boldsymbol{\Psi}_{\mathcal{S}_{K^u}^u,\mathcal{F}^u}, \boldsymbol{\Psi}_{\mathcal{S}_1^d,\mathcal{F}^d}, ..., \boldsymbol{\Psi}_{\mathcal{S}_{K^d}^d,\mathcal{F}^d}, \overline{\mathbf{U}}\Big],$$
$$(15)$$

where M in (13) is equal to $K^d + K^u + K^h$, $K^h = \dim(\mathbf{ker}(\mathbf{L}))$. The size of the dictionary built in (13) can be controlled either tuning K^d and K^u, or by taking just the top $\widetilde{K}$ Slepians per each pair of concentration sets. Clearly, the maximum possible number of Slepians depends on the chosen sets $\{\mathcal{S},\mathcal{F}\}$, because they affect the rank of the operator $\mathbf{B}_{\mathcal{F}}\mathbf{C}_{\mathcal{S}}\mathbf{B}_{\mathcal{F}}$ in (12). Therefore, for a given $\widetilde{K}$ and pair of concentration sets $\{\mathcal{S},\mathcal{F}\}$, the number of topological Slepians will be $\min\{\widetilde{K}, \text{rank}\{\mathbf{B}_{\mathcal{F}}\mathbf{C}_{\mathcal{S}}\mathbf{B}_{\mathcal{F}}\}\}$.

Frame bounds. The dictionary $\mathbf{D}_{\mathcal{C}} \in \mathbb{R}^{E \times M}$ in (15) is generally overcomplete, being composed by a set of $M > E$ non-orthogonal atoms. It is then of interest to investigate frame bounds on $\mathbf{D}_{\mathcal{C}}$. For a Hilbert space $\mathcal{H}$, a dictionary of vectors $\mathbf{D}_{\mathcal{C}}$ with at most countably many elements is an (A, B)-frame if [24]:

$$A||\mathbf{v}||_2^2 \leq \sum_{\boldsymbol{\psi} \in \mathbf{D}_{\mathcal{C}}} |<\boldsymbol{\psi}, \mathbf{v}>|^2 \leq B||\mathbf{v}||_2^2, \qquad (16)$$

for all $\mathbf{v} \in \mathcal{H}$, where A and B are called frame bounds, with $0 \leq A \leq B < \infty$. If $A = B$, we say that $\mathbf{D}_{\mathcal{C}}$ is a tight frame; if $A = 0$, $\mathbf{D}_{\mathcal{C}}$ is said to be a degenerate frame. Let $\boldsymbol{\Psi}^u = \{\boldsymbol{\Psi}_{\mathcal{S}_i^u,\mathcal{F}^u}\}_{i=1}^{K^u}$ and $\boldsymbol{\Psi}^d = \{\boldsymbol{\Psi}_{\mathcal{S}_i^d,\mathcal{F}^d}\}_{i=1}^{K^d}$ be the matrices collecting all the upper and lower Slepians, respectively. The following Theorem shows that, under mild conditions on the choice of the edge concentration sets, a Slepian dictionary as in (13) forms an non-degenarate (A, B)-frame.

Theorem 2. Let $\mathbf{D}_{\mathcal{C}}$ be a dictionary of topological Slepians as in (15). Let the following properties hold:
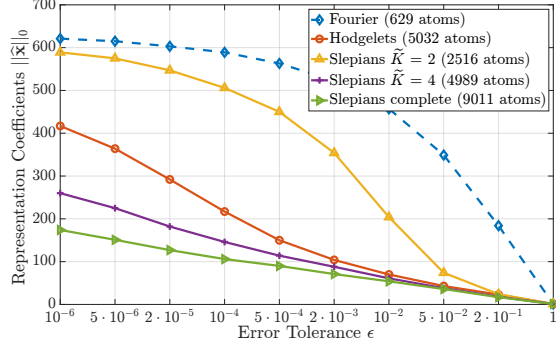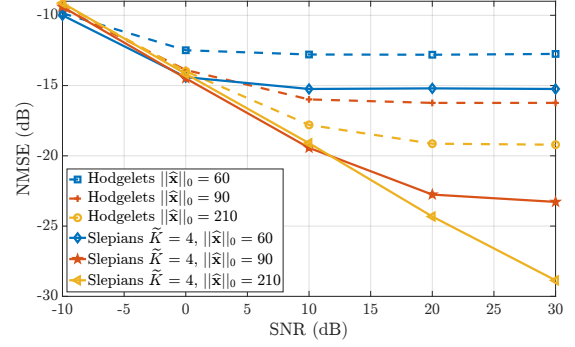
Fig. 2: Sparsity vs error tolerance.

Fig. 3: NMSE vs SNR, for different denoising strategies.

A1 (Lower completeness): The lower localization sets $\{\mathcal{S}_i^d\}_{i=1}^{K^d}$ are chosen such that $\text{colspan}\{\boldsymbol{\Psi}^d\} = \mathbf{im}(\mathbf{L}^d)$.

A2 (Upper completeness): The upper localization sets $\{\mathcal{S}_i^u\}_{i=1}^{K^d}$ are chosen such that $\text{colspan}\{\boldsymbol{\Psi}^u\} = \mathbf{im}(\mathbf{L}^u)$.

Then, $\mathbf{D}_\mathcal{C}$ is a non-degenerate (A, B)-frame for $\mathbb{R}^E$ with $A = \lambda_{min}(\sum_{\boldsymbol{\psi} \in \mathbf{D}_\mathcal{C}} \boldsymbol{\psi}\boldsymbol{\psi}^T) > 0$ and $B = K^d + K^u + K^h$.

Proof. The proof is omitted due to space limitations. ∎

In Theorem 2, A1 requires that the set of lower Slepians spans the gradient space, i.e., $\mathbf{im}(\mathbf{L}^d)$. Similarly, A2 assumes that the set of upper Slepians spans the curl space, i.e., $\mathbf{im}(\mathbf{L}^u)$. These conditions are sufficient for the claim of the theorem, and subsume some necessary conditions for the choice of the localization sets. The first one is related to coverage, i.e., the union of the localization sets defining the lower and upper Slepians must cover the whole edge set. Mathematically, we need $\bigcup_{i=1}^{K^d} \mathcal{S}_i^d = \bigcup_{i=1}^{K^u} \mathcal{S}_i^u = \mathcal{E}$. The second condition is related to the number of elements in the dictionary that must satisy $K^d \geq |\mathcal{F}^d|$ and $K^u \geq |\mathcal{F}^u|$, which is necessary to find a set of lower and upper Slepians that span $|\mathcal{F}^d|$ and $|\mathcal{F}^u|$ dimensional subspaces, i.e., the gradient and the curl spaces, respectively. Interestingly, the conditions of Theorem 2 can be easily satisfied using very simple heuristic procedures for the selection of the localization sets $\{\mathcal{S}_i^d\}_{i=1}^{K^d}$ and $\{\mathcal{S}_j^u\}_{j=1}^{K^u}$. For instance, an easy and low-complexity example is building a dictionary by choosing the upper and lower edge concentration sets as the 1-hop upper and lower neighborhoods of each edge including the edge itself, respectively. In the following numerical results, we will exploit this simple procedure, but the choice of the localization sets is an interesting open problem that could be addressed resorting to hierarchical higher order spectral clustering techniques, with the objective of finding tighter frames.

5. NUMERICAL RESULTS

We assess the performance of topological Slepians dictionaries on two tasks, i.e., sparse signal representation, and edge flow denoising. We rely on the experimental setting of [25], considering the localized vector field on $[-2, 2]^2$ given by $F(x, y) = [\cos(x + y), \sin(x - y)]\mathbb{I}\big((x, y) \in B_1 \cup B_2\big)$, where $\mathbb{I}$ is the indicator function, B_1 and B_2 are closed balls of radius 0.7 centered at $(\pm\pi/4, \pm\pi/4)$, respectively. The vector field is then discretized with a hexagonal grid, and a simplicial complex is built ($V = 225$, $E = 629$, $T = 405$) by treating each hexagon as a node, with edges for each pair of hexagons that share a side, and triangles for each set of three hexagons that share a corner. Finally, the vector field is converted to an edge signal $\mathbf{x} \in \mathbb{R}^E$ by taking the flow on each edge to be the total flow perpendicular to the corresponding side between hexagons; $\mathbf{x}$ is normalized

to have unit norm. The dictionary of topological Slepians is built as in (15), considering $2E$ edge concentration sets, defined as the upper and the lower 1-hop neighborhood of each edge, in order to satisfy the conditions of Theorem 2. No harmonic component is present.

Sparse signal representation - We first evaluate how sparse is the representation of $\mathbf{x}$ on the proposed dictionary. We employ orthogonal matching pursuit (OMP), which seeks the sparsest linear combination of the Slepians in the dictionary by greedily solving [34]:

$$\widehat{\mathbf{x}} = \arg\min ||\widehat{\mathbf{x}}||_0 \quad \text{subject to} \quad ||\mathbf{x} - \mathbf{D}_\mathcal{C}\widehat{\mathbf{x}}||^2 \leq \epsilon. \quad (17)$$

In Fig. 2, we illustrate the ℓ_0 norm of the obtained $\widehat{\mathbf{x}}$ versus the error tolerance ϵ, comparing the Fourier basis, the Hodgelets [25], and the topological Slepians dictionaries obtained for some values of $\widetilde{K}$. As we can see from Fig. 2, the proposed Slepians dictionary shows superior performance with respect to the Hodgelets [25], being able to find sparser representations even for smaller dictionaries, i.e., for $\widetilde{K} = 4$. The complete Slepians dictionary leads to the best performance, at the cost of a larger dictionary size and complexity.

Edge flow denoising - As a second experiment, we evaluate the effectiveness of topological Slepians diactionaries on a task of edge flow denoising. In particular, we take the same edge signal $\mathbf{x}$ of the previous experiment and we add independent Gaussian noise to it, obtaining $\widetilde{\mathbf{x}} = \mathbf{x} + \mathbf{n}$, where $\mathbf{n} \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$. Now, we exploit again OMP, but switching the objective function and the constraint of (17), so as to minimizee the mean squared error, under a sparsity constraint on the ℓ_0-norm of $\widehat{\mathbf{x}}$, with the goal of evaluating how good is the obtained representation $\mathbf{D}_\mathcal{C}\widehat{\mathbf{x}}$ in estimating the clean signal $\mathbf{x}$. In Fig. 3, we show the Normalized Mean Square error $||\mathbf{x} - \mathbf{D}_\mathcal{C}\widehat{\mathbf{x}}||^2$ versus the Signal to Noise Ratio (SNR), averaged over 100 independent simulations, comparing Hodgelets and topological Slepians dictionary (with $\widetilde{K} = 4$) obtained for three values of $||\widehat{\mathbf{x}}||_0$; the SNR is defined as SNR $= ||\mathbf{x}||^2/(\sigma^2 \cdot E) = 1/(\sigma^2 \cdot E)$. From Fig. 3 it is evident that topological Slepians leads to better estimation performance, for any value of SNR and $||\widehat{\mathbf{x}}||_0$.

6. CONCLUSIONS

In this paper we introduced topological Slepians, a novel class of signals that are maximally localized over a simplicial complex, while being perfectly localized onto a dual (frequency) domain. Hinging on Hodge decomposition and exploiting the structure of high-order Laplacians, we have proposed a principled method to build dictionaries of topological Slepians and we derived their frame bounds. Finally, we assessed their effectiveness in sparse signal representation and edge flow denoising when compared to state of the art approaches available in the literature.

7. REFERENCES

[1] 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.

[2] 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.

[3] S. Klamt, U.-U. Haus, and F. Theis, "Hypergraphs and cellular networks," *PLoS Comput. Biol.*, vol. 5, no. 5, 2009.

[4] R. Lambiotte, M. Rosvall, and I. Scholtes, "From networks to optimal higher-order models of complex systems," *Nature physics*, vol. 15, no. 4, pp. 313–320, 2019.

[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," *Journal of computational neuroscience*, 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. Lear.*, 2006, pp. 17–24.

[10] S. Barbarossa and S. Sardellitti, "Topological signal processing over simplicial complexes," *IEEE Trans. on Signal Processing*, vol. 68, pp. 2992–3007, 2020.

[11] M. T. Schaub, Y. Zhu, J.B. Seby, T. M. Roddenberry, and S. Segarra, "Signal processing on higher-order networks: Livin'on the edge... and beyond," *Signal Processing*, vol. 187, pp. 108149, 2021.

[12] Maosheng Yang, Elvin Isufi, Michael T. Schaub, and Geert Leus, "Simplicial convolutional filters," 2022.

[13] M. T. Schaub, A. R. Benson, P. Horn, G. Lippner, and A. Jadbabaie, "Random walks on simplicial complexes and the normalized hodge 1-laplacian," *SIAM Review*, vol. 62, no. 2, pp. 353–391, Jan 2020.

[14] Feng Ji, Giacomo Kahn, and Wee Peng Tay, "Signal processing on simplicial complexes with vertex signals," *IEEE Access*, vol. 10, pp. 41889–41901, 2022.

[15] M. Coutino, G. V. Karanikolas, and G. B. Leus, G.and Giannakis, "Self-driven graph volterra models for higher-order link prediction," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2020, pp. 3887–3891.

[16] Stefania Sardellitti, Sergio Barbarossa, and Lucia Testa, "Topological signal processing over cell complexes," in *2021 55th Asilomar Conference on Signals, Systems, and Computers*, 2021, pp. 1558–1562.

[17] T. Mitchell Roddenberry, Michael T. Schaub, and Mustafa Hajij, "Signal processing on cell complexes," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 8852–8856.

[18] C. Bodnar, F. Frasca, Y. Guang Wang, N. Otter, G. Montufar, P. Liò, and M. M. Bronstein, "Weisfeiler and Lehman go topological: Message passing simplicial networks," in *ICLR 2021 Workshop on Geometrical and Topological Representation Learning*, 2021.

[19] E. Isufi, F. Gama, and A. Ribeiro, "Edgenets: Edge varying graph neural networks," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2021.

[20] S. Ebli, M. Defferrard, and G. Spreemann, "Simplicial neural networks," in *NeurIPS 2020 Workshop on Topological Data Analysis and Beyond*, 2020.

[21] T. M. Roddenberry and S. Segarra, "Hodgenet: Graph neural networks for edge data," *Computing Research Repository (CoRR)*, vol. abs/1912.02354, 2019.

[22] L. Giusti, C. Battiloro, P. Di Lorenzo, S. Sardellitti, and S. Barbarossa, "Simplicial attention neural networks," arXiv:2203.07485, 2022.

[23] Lorenzo Giusti, Claudio Battiloro, Lucia Testa, Paolo Di Lorenzo, Stefania Sardellitti, and Sergio Barbarossa, "Cell attention networks," *arXiv preprint arXiv:2209.08179*, 2022.

[24] O. Rioul and M. Vetterli, "Wavelets and signal processing," *IEEE Signal Proc. Mag.*, vol. 8, no. 4, pp. 14–38, 1991.

[25] T. Mitchell Roddenberry, Florian Frantzen, Michael T. Schaub, and Santiago Segarra, "Hodgelets: Localized spectral representations of flows on simplicial complexes," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 5922–5926.

[26] David I Shuman, Christoph Wiesmeyr, Nicki Holighaus, and Pierre Vandergheynst, "Spectrum-adapted tight graph wavelet and vertex-frequency frames," *IEEE Transactions on Signal Processing*, vol. 63, no. 16, pp. 4223–4235, 2015.

[27] D. Slepian and H. O. Pollak, "Prolate spheroidal wave functions, fourier analysis and uncertainty — i," *Bell System Technical Journal*, vol. 40, no. 1, pp. 43–63, 1961.

[28] Mikhail Tsitsvero, Sergio Barbarossa, and Paolo Di Lorenzo, "Signals on graphs: Uncertainty principle and sampling," *IEEE Transactions on Signal Processing*, vol. 64, no. 18, pp. 4845–4860, 2016.

[29] Dimitri Van De Ville, Robin Demesmaeker, and Maria Giulia Preti, "Guiding network analysis using graph slepians: an illustration for the c. elegans connectome," in *Optical Engineering + Applications*, 2017.

[30] L. Lim, "Hodge Laplacians on graphs," *Siam Review*, vol. 62, no. 3, pp. 685–715, 2020.

[31] M. Yang, E. Isufi, M. T. Schaub, and G. Leus, "Finite impulse response filters for simplicial complexes," *arXiv preprint arXiv:2103.12587*, 2021.

[32] T. E. Goldberg, "Combinatorial Laplacians of simplicial complexes," *Senior Thesis, Bard College*, 2002.

[33] M. Yang, E. Isufi, and G. Leus, "Simplicial convolutional neural networks," *arXiv preprint arXiv:2110.02585*, 2021.

[34] Joel A. Tropp and Anna C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. on Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.