Hodge-Aware Matched Subspace Detectors

Liu, Chengen; Isufi, Elvin

DOI
10.23919/EUSIPCO63174.2024.10714960

Publication date
2024

Document Version
Final published version

Published in
32nd European Signal Processing Conference, EUSIPCO 2024 - Proceedings

Hodge-Aware Matched Subspace Detectors

Chengen Liu and Elvin Isufi
Delft University of Technology, Delft, The Netherlands
Email:{C.Liu-15; E.Isufi-1}@tudelft.nl

***Abstract*—This paper introduces a hypothesis testing problem to detect whether a noisy simplicial signal lives in some specific Hodge subspaces or not. This is of particular relevance for edge flows in a network since they exhibit, under normal circumstances, different properties in Hodge decomposition. For example, a traffic flow in a road network is often conservative and that can be localized in a particular Hodge subspace. We propose two Neyman-Pearson optimal detectors for this task: the Simplicial Hodge Detector (SHD) and the Constrained Simplicial Hodge Detector (CSHD). They compare the energy of the simplicial embeddings in different Hodge subspaces and distinguish between the two hypotheses. The SHD utilizes the maximum likelihood estimation, while CSHD incorporates signal prior information to estimate the simplicial embeddings. These detectors are validated through numerical simulations on both real-world and synthetic data, indicating great potential in practical applications.**

***Index Terms*—Signal processing over higher-order networks, detection theory, topological signal processing.**

I. Introduction

Data with higher-order structures contain more intrisic information than graphs or Euclidean structured data such as images [1] [2] [3]. These structures include simplicial complexes, cell complexes, and hypergraphs [2] [4] [5] and encode multi-way relations while graphs are limited only to pair-wise relations. For example, flows can be modeled as signals supported on a set of edges [6] [7], thereby exploiting concepts in a higher-order network framework. Tools in signal processing and machine learning have been developed to process signals defined on higher-order networks such as convolutional and trend filters [8] [9], neural networks [10], Fourier analysis [11], and autoregressive model [12].

Central to these processing techniques is their algebraic representation of the underlying structure via the Hodge Laplacian [11] [13] [14]. The Hodge decomposition of this Laplacian states that any simplicial signal (data defined over simplices of a given level) comprises the sum of three components which are the gradient, the curl, and the harmonic component [8] [15]. These components live in orthogonal subspaces and carry specific interpretations. For example, for edge flows, for example, the gradient component is induced by the difference of signals defined on the nodes, which is similar to the relationship between electric potential and current. The curl component is the edge flow around triangles, which can be regarded as the local circulations. And the harmonic component is the flow-conservative part [8]. The projections of the signal components in the specific subspaces yield their Hodge embeddings, which are the bases for analyzing simplicial signals.

This paper is supported by the Dutch Grant GraSPA (No. 19497) financed by the Netherlands Organization for Scientific Research (NWO). Chengen Liu receives funding from the China Scholarship Council.

Many real-world signals with a higher-order structure possess distinctive properties that can be characterized by their Hodge embeddings. Two such properties are the curl-free and the divergence-free property [16]. Curl-free means that the sum of the edge flows circulating around triangles is zero, and divergence-free means the total amount of edge flows entering a node equals the flow leaving this node [8]. For instance, in a currency exchange market, the arbitrage free condition indicates that the logarithm exchange rate flow is curl-free, which means that profit cannot be made by exchanging between different currency pairs repeatedly [6]. Likewise, in a road network, traffic flows are often divergence-free since the cars on roads are conservative [8] [17].

When abnormalities occur, such as estimation error or data missing [6] [17], the above signal properties are no longer satisfied. Therefore, it is meaningful to detect such cases from a few measurements in a mathematically tractable way. This falls naturally under a matched subspace detection (MSD) perspective (the signal is in one particular subspace, e.g., gradient or not) [18]. While machine learning classifiers can be used, they often lack tractability or require a large amount of data. Therefore, we rely on MSD principles to build optimal and mathematically tractable detectors that can act as classifiers in a limited data regime. Related works on MSD have been proposed in the graph domain, such as the bandlimited signal detector in [19] and the blind graph topology detector in [20] but they do not apply to data on higher-order networks.

We propose a hypothesis test problem to detect whether a simplicial signal lives in a specific Hodge subspace or not. We exploit the Hodge subspace projections of the different components [8] and due to their orthogonality the embeddings have different shapes and probability distributions. Then, we consider a generalized likelihood ratio test to estimate the signals from noisy observations [21] and make the following contributions.

- We develop a MSD theory for simplicial signals based on the properties of the different Hodge subspaces.
- Inspired by the MSD for graphs [19] and [20], we propose two simplicial detectors, the Simplicial Hodge Detector (SHD) and the Constrained Simplicial Hodge Detector (CSHD). They are energy-type detectors which consider the projection of the signals in Hodge subspaces. SHD exploits the maximum likelihood principle to estimate the embedded signal, while CSHD considers the prior

information and exploits regularized maximum likelihood estimators.
- We conduct experiments on real-world and synthetic datasets to verify the effectiveness of the proposed detectors.

II. Preliminary

A. Simplicial Complexes

Given a set of vertices $\mathcal{V}$, a k-simplex $\mathcal{S}^k$ is a subset of $\mathcal{V}$ containing $k+1$ elements. A simplicial complex of order K, $\mathcal{P}^K$, is a finite collection of k-simplices $\mathcal{S}^k$ for $k = 0, 1, ..., K$ satisfying the inclusion property: for any $\mathcal{S}^k \in \mathcal{P}^K$, all of its subsets $\mathcal{S}^{k-1} \subset \mathcal{S}^k$ are included in the simplicial complex, i.e., $\mathcal{S}^{k-1} \in \mathcal{P}^K$. The number of k-simplices in $\mathcal{P}^K$ is N_k. By considering the geometric embedding of the simplicial complex into the Euclidean space, a node can be seen as a 0-simplex, an edge as a 1-simplex, and a (filled) triangle as a 2-simplex; see Figure 1. A graph is therefore a simplicial complex of order $K = 1$ including nodes and edges.

The adjacencies between different simplices can be described by the incidence matrices $\mathbf{B}_k \in \mathbb{R}^{N_{k-1} \times N_k}$ which represent the relationship between (k-1)-simplices and k-simplices [11]. These incidence matrices can be used to build the *Hodge Laplacian* matrices that can represent the structure of a simplicial complex

$$\begin{aligned} \mathbf{L}_0 &= \mathbf{B}_1 \mathbf{B}_1^\top \\ \mathbf{L}_k &= \mathbf{B}_k^\top \mathbf{B}_k + \mathbf{B}_{k+1} \mathbf{B}_{k+1}^\top, k = 1, \ldots, K-1 \\ \mathbf{L}_K &= \mathbf{B}_K^\top \mathbf{B}_K. \end{aligned} \tag{1}$$

Except for $\mathbf{L}_0$ and $\mathbf{L}_K$, the other intermediate Laplacians $\mathbf{L}_k$ consist of the *lower Laplacian* $\mathbf{L}_{k,\ell} = \mathbf{B}_k^\top \mathbf{B}_k$ and the *upper Laplacian* $\mathbf{L}_{k,u} = \mathbf{B}_{k+1} \mathbf{B}_{k+1}^\top$. They represent the lower-adjacencies (e.g., how two edges are adjacent via a common node), and the upper-adjacencies (e.g., how two edges are adjacent by being the faces of the same triangle) of k-simplices respectively.

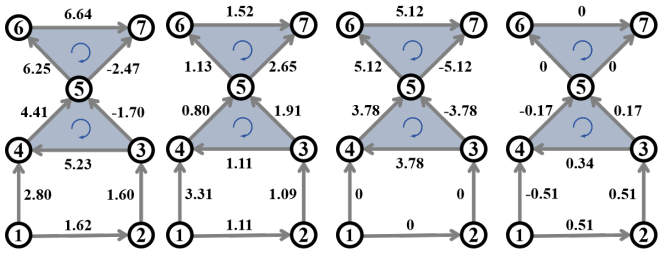
A $k-$simplicial signal $\mathbf{s}^k = \left[s_1^k, \ldots, s_{N_k}^k\right]^\top \in \mathbb{R}^{N_k}$ is supported on k-simplices where entry s_i^k corresponds to the ith k-simplex [11]. For simplicity of notation, hereafter, $0-$, $1-$ and $2-$simplicial signal are denoted as $\mathbf{v}$, $\mathbf{f}$ and $\mathbf{t}$. They represent node, edge, and triangle signals, respectively. A simplex can have two orientations and for processing purposes. We follow the lexicographical ordering of the vertices to define the reference orientation. If the value of the signal is positive, the set orientations are consistent with the real situation. If it is negative, the set orientations are opposite.

B. Hodge Decomposition

The *Hodge decomposition* states that the space of the $k-$simplicial signal $\mathbb{R}^{N_k}$ can be decomposed into three orthogonal subspaces

$$\mathbb{R}^{N_k} \equiv \operatorname{span}\left(\mathbf{B}_k^\top\right) \oplus \operatorname{kernel}\left(\mathbf{L}_k\right) \oplus \operatorname{span}\left(\mathbf{B}_{k+1}\right) \tag{2}$$

where $\oplus$ is the direct sum operator. In this paper, we focus on edge flows due to their wider applicability, i.e., $1-$simplicial signal. Then, $\operatorname{span}\left(\mathbf{B}_1^\top\right)$, $\operatorname{span}\left(\mathbf{B}_2\right)$ and $\operatorname{kernel}\left(\mathbf{L}_1\right)$ are the gradient, the curl, and the harmonic subspace with dimension



(a) Edge flow $\mathbf{f}$. (b) $\mathbf{f}_{\mathrm{G}}$. (c) $\mathbf{f}_{\mathrm{C}}$. (d) $\mathbf{f}_{\mathrm{H}}$.

Fig. 1: Hodge decomposition of the simplicial signal for a simplicial complex of order two. The edge flow can be decomposed into three different components: the gradient $\mathbf{f}_G$, the curl $\mathbf{f}_C$ and the harmonic component $\mathbf{f}_H$.

N_G, N_C and N_H, respectively. To reveal the relationships between the different Hodge subspaces, we consider the eigen-decomposition of the first Hodge Laplacian

$$\mathbf{L}_1 = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^\top \tag{3}$$

where $\mathbf{U} \in \mathbb{R}^{N_1 \times N_1}$ is an orthonormal matrix that collects the eigenvectors and $\mathbf{\Lambda} = \operatorname{diag}(\lambda_1, ...\lambda_{N_1}) \in \mathbb{R}^{N_1 \times N_1}$ is a diagnal matrix that collects the corresponding eigenvalues. The columns of $\mathbf{U}$ can be rearranged into $[\mathbf{U}_{\mathrm{G}}\ \mathbf{U}_{\mathrm{C}}\ \mathbf{U}_{\mathrm{H}}]$ with the following explanation.

- *Gradient eigenvectors*: $\mathbf{U}_{\mathrm{G}} \in \mathbb{R}^{N_1 \times N_G}$ collects the eigenvectors of $\mathbf{L}_{1,\ell}$ corresponding to the eigenvalues $\lambda_{\mathrm{G},i} > 0$ and span the gradient space $\operatorname{span}\left(\mathbf{B}_1^\top\right)$.
- *Curl eigenvectors*: $\mathbf{U}_{\mathrm{C}} \in \mathbb{R}^{N_1 \times N_C}$ collects the eigenvectors of $\mathbf{L}_{1,u}$ corresponding to the eigenvalues $\lambda_{\mathrm{C},i} > 0$ and span the curl space $\operatorname{span}\left(\mathbf{B}_2\right)$.
- *Harmonic eigenvectors*: $\mathbf{U}_{\mathrm{H}} \in \mathbb{R}^{N_1 \times N_H}$ collects the eigenvectors of $\mathbf{L}_1$ corresponding to the zero eigenvalues $\lambda_{\mathrm{H},i} = 0$ and span the harmonic space $\operatorname{kernel}(\mathbf{L}_1)$.

The Hodge decomposition also implies that for any $1-$simplicial signal $\mathbf{f}$, there exist three signals of orders 0, 1, and 2 so that we can decompose the $1-$simplicial signal as

$$\mathbf{f} = \mathbf{B}_1^\top \mathbf{v} + \mathbf{f}_{\mathrm{H}} + \mathbf{B}_2 \mathbf{t}. \tag{4}$$

Thus, any edge flow can be written as a sum of three flows $\mathbf{f} = \mathbf{f}_{\mathrm{G}} + \mathbf{f}_{\mathrm{C}} + \mathbf{f}_{\mathrm{H}}$ (see also Fig. 1) with the following explanation:

- *Gradient component*: $\mathbf{f}_{\mathrm{G}} = \mathbf{B}_1^\top \mathbf{v} \in \operatorname{span}(\mathbf{B}_1^\top)$ is a flow in the gradient space and the gradient operator $\mathbf{B}_1^\top$ maps node signals $\mathbf{v}$ into edge signals $\mathbf{f}$. It is induced by taking the difference between the node pair signal values. Eigenvectors $\mathbf{U}_{\mathrm{G}}$ span the gradient space and the corresponding gradient embedding is $\hat{\mathbf{f}}_{\mathrm{G}} = \mathbf{U}_{\mathrm{G}}^\top \mathbf{f} = \mathbf{U}_{\mathrm{G}}^\top \mathbf{f}_{\mathrm{G}} \in \mathbb{R}^{N_G}$.
- *Curl component*: $\mathbf{f}_C = \mathbf{B}_2 \mathbf{t} \in \operatorname{span}(\mathbf{B}_2)$ is a flow in the curl space and the curl adjoint operator $\mathbf{B}_2$ maps triangle signals $\mathbf{t}$ into edge signals $\mathbf{f}$. It is the edge flow that circulates around a triangle. Eigenvectors $\mathbf{U}_{\mathrm{C}}$ span the curl space and the corresponding curl embedding is $\hat{\mathbf{f}}_{\mathrm{C}} = \mathbf{U}_{\mathrm{C}}^\top \mathbf{f} = \mathbf{U}_{\mathrm{C}}^\top \mathbf{f}_{\mathrm{C}} \in \mathbb{R}^{N_C}$.
- *Harmonic component*: $\mathbf{f}_{\mathrm{H}} \in \operatorname{kernel}(\mathbf{L}_1)$ is the flow in the harmonic space $\operatorname{kernel}\left(\mathbf{L}_1\right)$ with dimension N_H that satisfies $\mathbf{L}_1 \mathbf{f}_{\mathrm{H}} = \mathbf{0}$. Eigenvectors $\mathbf{U}_{\mathrm{H}}$ span the harmonic

space and the corresponding harmonic embedding is $\hat{\mathbf{f}}_{\mathrm{H}} = \mathbf{U}_{\mathrm{H}}^{\top}\mathbf{f} = \mathbf{U}_{\mathrm{H}}^{\top}\mathbf{f}_{\mathrm{H}} \in \mathbb{R}^{N_H}$.

Simplicial embeddings indicate that the gradient eigenvectors $\mathbf{U}_{\mathrm{G}}$ are orthogonal to the curl and harmonic component $\mathbf{f}_{\mathrm{C}}$ and $\mathbf{f}_{\mathrm{H}}$. This also holds similarly for the curl and the harmonic eigenvectors. Based on this orthogonality, in the next section, we propose the Hodge subspaces detectors. These three components exhibit two diverse properties that are seen in real signals:

- *Curl-free*: $\mathrm{curl}(\mathbf{f}) = \mathbf{B}_2^{\top}\mathbf{f}$ is the curl operator which measures the curl of an edge flow. Its *i*th element represents the aggregate flow across all edges that constitute the *i*th triangle. If $\mathrm{curl}(\mathbf{f}) = \mathbf{0}$, $\mathbf{f}$ is a curl-free flow. By definition, the gradient and harmonic flow are curl-free. For example, an ideal currency exchange rate flow is curl-free [6].
- *Divergence-free*: $\mathrm{div}(\mathbf{f}) = \mathbf{B}_1\mathbf{f}$ is the divergence operator which measures the divergence of an edge flow. Its *i*th element represents the total flow passing through the *i*th node. If all the elements in $\mathrm{div}(\mathbf{f})$ are zero, $\mathbf{f}$ is a divergence-free edge flow. By definition, the curl and the harmonic component are divergence-free. The traffic flow tends to be divergence free as the number of the cars in a road network is conservative [8].

III. Hodge Subspace Detectors

Real-world signals belong to one or more of the Hodge subspaces. However, under noisy observations or anomalies, their projections will not be located only in these subspaces. Therefore, detecting such a signal subspace spillage is of interest to reveal anomalies in the data. When no anomalies or noise is present, this detection is straightforward. However, that becomes challenging under noisy observations. Our objective is to detect whether a noisy edge flow signal belongs to a certain Hodge subspace or not. Specifically, let $\mathbf{x} \in \mathbb{R}^{N_1}$ denote the real edge flow and $\mathbf{f}$ the corresponding corrupted flow with some zero-mean Gaussian noise $\mathbf{n} \sim \mathcal{N}\left(\mathbf{0}_{N_1}, \sigma^2\mathbf{I}_{N_1}\right)$. A hypothesis test for this problem can be formulated as

$$\begin{aligned} \mathcal{H}_0: &\quad \mathbf{x} \text{ lives in some specific Hodge subspaces} \\ \mathcal{H}_1: &\quad \mathbf{x} \text{ does not live in these Hodge subspaces.} \end{aligned} \tag{5}$$

If flow $\mathbf{x}$ lives in some specific subspace, we consider that subspace is spanned by the eigenvectors

$$\mathbf{U}_{\Delta} \in \{\mathbf{U}_{\mathrm{G}}, \mathbf{U}_{\mathrm{C}}, \mathbf{U}_{\mathrm{H}}, [\mathbf{U}_{\mathrm{G}}, \mathbf{U}_{\mathrm{C}}], [\mathbf{U}_{\mathrm{G}}, \mathbf{U}_{\mathrm{H}}], [\mathbf{U}_{\mathrm{C}}, \mathbf{U}_{\mathrm{H}}]\}. \tag{6}$$

Notice that eigenvectors $\mathbf{U}_{\Delta}$ may be a union of two different conventional subspace eigenvevectors such as $[\mathbf{U}_{\mathrm{G}}, \mathbf{U}_{\mathrm{H}}]$ if the signal lives in two different subspaces e.g., a curl-free signal. This allows writing the flow $\mathbf{x}$ as $\mathbf{x} = \mathbf{U}_{\Delta}\hat{\mathbf{x}}_{\Delta}$. Likewise, let us also consider the complement (orthogonal) eigenvectors to $\mathbf{U}_{\Delta}$ which is

$$\mathbf{U}_{\overline{\Delta}} \in \{[\mathbf{U}_{\mathrm{C}}, \mathbf{U}_{\mathrm{H}}], [\mathbf{U}_{\mathrm{G}}, \mathbf{U}_{\mathrm{H}}], [\mathbf{U}_{\mathrm{G}}, \mathbf{U}_{\mathrm{C}}], \mathbf{U}_{\mathrm{H}}, \mathbf{U}_{\mathrm{C}}, \mathbf{U}_{\mathrm{G}}\}. \tag{7}$$

Then, the projection of $\mathbf{f}$ in the complement subspace is $\hat{\mathbf{f}}_{\overline{\Delta}} = \mathbf{U}_{\overline{\Delta}}^{\top}\mathbf{x} + \mathbf{U}_{\overline{\Delta}}^{\top}\mathbf{n} = \hat{\mathbf{x}}_{\overline{\Delta}} + \hat{\mathbf{n}}_{\overline{\Delta}}$. The projections $\hat{\mathbf{x}}_{\overline{\Delta}}$ and $\hat{\mathbf{n}}_{\overline{\Delta}}$ represent the clean signal and noise in the different Hodge subspaces, respectively. The projected noise satisfies $\hat{\mathbf{n}}_{\overline{\Delta}} \sim \mathcal{N}\left(\mathbf{0}_{N_{\overline{\Delta}}}, \sigma^2\mathbf{I}_{N_{\overline{\Delta}}}\right)$. Under hypothesis $\mathcal{H}_0$, if $\mathbf{x}$ lives in the subspaces spanned by $\mathbf{U}_{\Delta}$, the projection $\mathbf{U}_{\overline{\Delta}}^{\top}\mathbf{x}$ is $\mathbf{0}$ due to the orthogonality between the eigenvectors. Thus, the projection of $\mathbf{f}$ in the complement subspace under $\mathcal{H}_0$ is only noise $\hat{\mathbf{n}}_{\overline{\Delta}}$. Therefore, the hypothesis test (5) can be reformulated as

$$\begin{aligned} \mathcal{H}_0: &\quad \hat{\mathbf{f}}_{\overline{\Delta}} = \hat{\mathbf{n}}_{\overline{\Delta}} \\ \mathcal{H}_1: &\quad \hat{\mathbf{f}}_{\overline{\Delta}} = \mathbf{U}_{\overline{\Delta}}^{\top}\mathbf{x} + \hat{\mathbf{n}}_{\overline{\Delta}} \end{aligned}. \tag{8}$$

Equation (8) is the classical problem of detecting a signal $\mathbf{U}_{\overline{\Delta}}^{\top}\mathbf{x}$ corrupted by noise [18]. The true signal is unknown and we have to estimate it. Next, we show how this problem is solved by different matched subspace detection methods.

A. Simplicial Hodge Detector (SHD)

To distinguish between $\mathcal{H}_0$ and $\mathcal{H}_1$, we consider the generalized likelihood ratio test (GLRT):

$$T(\hat{\mathbf{f}}_{\overline{\Delta}}) = \frac{p\left(\hat{\mathbf{f}}_{\overline{\Delta}}; \hat{\mathbf{x}}^*_{\overline{\Delta}1}, \mathcal{H}_1\right)}{p\left(\hat{\mathbf{f}}_{\overline{\Delta}}; \hat{\mathbf{x}}^*_{\overline{\Delta}0}, \mathcal{H}_0\right)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma. \tag{9}$$

Here $p\left(\hat{\mathbf{f}}_{\overline{\Delta}}; \hat{\mathbf{x}}^*_{\overline{\Delta}j}, \mathcal{H}_j\right)$ is the probability density function and $\hat{\mathbf{x}}^*_{\overline{\Delta}j}$ is the maximum likelihood estimator (MLE) of the true flow embedding under hypothesis $\mathcal{H}_j, j \in \{0, 1\}$. Scalar γ is the decision threshold. Since the noise is zero-mean Gaussian, the probability density function is

$$p\left(\hat{\mathbf{f}}_{\overline{\Delta}}; \hat{\mathbf{x}}^*_{\overline{\Delta}}, \mathcal{H}_j\right) = \left(2\pi\sigma^2\right)^{\frac{-N_{\overline{\Delta}}}{2}} \exp\left\{\frac{-\|\hat{\mathbf{f}}_{\overline{\Delta}} - \hat{\mathbf{x}}^*_{\overline{\Delta}}\|_2^2}{2\sigma^2}\right\}. \tag{10}$$

Under hypothesis $\mathcal{H}_1$, the MLE maximizes the probability density function, thus, $\hat{\mathbf{x}}^*_{\overline{\Delta}1} = \hat{\mathbf{f}}_{\overline{\Delta}}$ while under hypothesis $\mathcal{H}_0$, $\hat{\mathbf{x}}^*_{\overline{\Delta}0} = \mathbf{0}$. Therefore, the SHD is

$$T_{\mathrm{SHD}}(\hat{\mathbf{f}}_{\overline{\Delta}}) = \|\hat{\mathbf{f}}_{\overline{\Delta}}\|_2^2/\sigma^2 \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma. \tag{11}$$

When $T_{\mathrm{SHD}}(\hat{\mathbf{f}}_{\overline{\Delta}})$ is larger than threshold γ, the detector decides $\mathcal{H}_1$. That is, the SHD decides $\mathcal{H}_1$ if the signal-to-noise ratio of the complementary simplicial embedding exceeds a threshold γ. The detector $T_{\mathrm{SHD}}(\hat{\mathbf{f}}_{\overline{\Delta}})$ has a Chi-square distribution

$$T_{\mathrm{SHD}}(\hat{\mathbf{f}}_{\overline{\Delta}}) \sim \begin{cases} \chi^2_{N_{\overline{\Delta}}} & \text{under } \mathcal{H}_0 \\ \chi^2_{N_{\overline{\Delta}}}(\delta) & \text{under } \mathcal{H}_1 \end{cases} \tag{12}$$

where $N_{\overline{\Delta}}$ are the degrees of freedom and δ is a noncentrality parameter satisfying $\delta = \|\hat{\mathbf{x}}_{\overline{\Delta}}\|_2^2/\sigma^2$. The performance of the SHD detector can be computed in closed-form [20]. Theoretically, the detection threshold γ is

$$\gamma = Q^{-1}_{\chi^2_{N_{\overline{\Delta}}}}(\mathrm{P_{FA}}) \tag{13}$$

where $Q_{\chi^2_{N_{\overline{\Delta}}}}(\cdot)$ is the right-tail probability function and $\mathrm{P_{FA}} \triangleq \Pr\{T_{\mathrm{SHD}}(\hat{\mathbf{f}}_{\overline{\Delta}}) > \gamma; \mathcal{H}_0\}$ is the probability of false alarm. The probability of detection is

$$\mathrm{P_D} \triangleq \Pr\{T_{\mathrm{SHD}}(\hat{\mathbf{f}}_{\overline{\Delta}}) > \gamma; \mathcal{H}_1\} = Q_{\chi^2_{N_{\Delta}}(\delta)}(\gamma). \tag{14}$$

B. Constrained Simplicial Hodge Detector (CSHD)

Simplicial embeddings could exhibit particular behaviors among the subspaces. For example, they could be sparse (i.e., bandlimited) or smooth. Exploiting this knowledge could improve the performance of a hypothesis test [19]. The prior knowledge can be utilized as a regularizer in the MLE problem with respect to the simplicial embeddings. By regularizing the maximum log-likelihood function, we have:

$$\hat{\mathbf{x}}_{\overline{\Delta}}^{*} = \arg\max_{\hat{\mathbf{x}}_{\overline{\Delta}}} \mathcal{L}\left(\hat{\mathbf{x}}_{\overline{\Delta}}\right) - \mu \cdot r(\hat{\mathbf{x}}_{\overline{\Delta}}) \tag{15}$$

where $\mathcal{L}(\cdot)$ is the log-likelihood function, $r(\hat{\mathbf{x}}_{\overline{\Delta}})$ is the regularizer on the simplicial embeddings and $\mu > 0$ is a constant. For instance, if the simplicial embedding $\hat{\mathbf{x}}_{\overline{\Delta}}$ is smooth and hence low-pass, we can set the regularizer as $r(\hat{\mathbf{x}}_{\overline{\Delta}}) = \|\mathbf{R}\hat{\mathbf{x}}_{\overline{\Delta}}\|_2^2$. Here $\mathbf{R} \in \mathbb{R}^{N_{\overline{\Delta}} \times N_{\overline{\Delta}}}$ is a diagonal matrix decreasing diagonal elements. This allows to control the shape of estimated embedding $\hat{\mathbf{x}}_{\overline{\Delta}}$ based on the low-pass prior. The embedding is then estimated by solving the regularized least squares problem

$$\arg\min_{\hat{\mathbf{x}}_{\overline{\Delta}}} \frac{1}{2\sigma^2}\|\hat{\mathbf{x}}_{\overline{\Delta}} - \hat{\mathbf{f}}_{\overline{\Delta}}\|_2^2 + \mu\|\mathbf{R}\hat{\mathbf{x}}_{\overline{\Delta}}\|_2^2 \tag{16}$$

with the closed-form constrained MLE solution

$$\hat{\mathbf{x}}_{\overline{\Delta}}^{*} = (\mathbf{I} + 2\mu\sigma^2\mathbf{R}^\top\mathbf{R})^{-1}\hat{\mathbf{f}}_{\overline{\Delta}}. \tag{17}$$

Plugging (17) instead of the MLE $\hat{\mathbf{x}}_{\overline{\Delta}1}^{*} = \hat{\mathbf{f}}_{\overline{\Delta}}$ into (10), we obain the CSHD

$$T_{\mathrm{CSHD}}(\hat{\mathbf{f}}_{\overline{\Delta}}) = (\|\hat{\mathbf{f}}_{\overline{\Delta}}\|_2^2 - \|\hat{\mathbf{f}}_{\overline{\Delta}} - \hat{\mathbf{x}}_{\overline{\Delta}}^{*}\|_2^2)/\sigma^2 \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma. \tag{18}$$

CSHD compares the nonmatched[1] signal-to-noise ratio with a threshold γ and decides $\mathcal{H}_1$ if it exceeds γ. The signal energy term $\|\hat{\mathbf{f}}_{\overline{\Delta}}\|_2^2$ is corrected by $\|\hat{\mathbf{f}}_{\overline{\Delta}} - \hat{\mathbf{x}}_{\overline{\Delta}}^{*}\|_2^2$ based on prior information. Its distribution is not Chi-square anymore and its theoretical analysis is challenging.

IV. Numerical Results

We verify the proposed detectors through numerical results on two real datasets: the Forex [6], and the Lastfm dataset [6], as well as a synthetic dataset from the Chicago road network [16]. Our results are averaged over 5×10^4 realizations.

A. Datasets

Forex [6]. This dataset contains the exchange rates between 25 different currencies that satisfy the arbitrage free condition. This currency exchange market can be described by a simplicial complex contains 25 nodes, 300 edges and 2300 triangles. The nodes represent different currencies, edges the exchange rate between two currencies, and all triangles are considered filled. If we take the logarithm of the exchange rates and model them as edge flows, the arbitrage free condition indicates that such flow is curl-free. In an ideal exchange rate scenario, the flow is a gradient flow [8]. The technical task is to detect whether an observed edge flow is a gradient flow.

Lastfm [6]. This dataset records the process of users switching artists while playing music. The transition process can be described by a simplicial complex with 657 nodes, 1997 edges, and 1276 triangles. The artists are modeled as nodes, the transitions between artists as edges, and the filled triangles represent the transitions forming a closed loop for any three artists. The edge flows depict the number of the transitions between different artists. If the user switches from artist A to B, then a unit is added on the edge between A and B. An edge flow modeled in this way is divergence-free since the user consistently moves on to different artists after one. The task is to detect whether an edge flow is divergence-free.

Chicago road network [16]. This is the Chicago transportation network which can be modeled by a simplicial complex with 546 nodes, 1088 edges and 112 triangles. The junctions are the nodes, the roads are the edges, and the areas enclosed by three roads are the triangles. Since the traffic flow can be regarded as divergence-free when there is no congestion [8], we generate synthetic curl flow on this simplicial complex. The task is to detect whether an observed edge flow is divergence-free or not.

B. Experimenal Setup

The experimental setup is summarized in TABLE I. For the Forex dataset, we consider the real gradient flow under hypothesis $\mathcal{H}_0$ and two flows under hypothesis $\mathcal{H}_1$ which are a synthetic curl flow and the sampled version of the gradient flow, respectively.

For the Lastfm, we consider the divergence-free user transition flows under $\mathcal{H}_0$ and two flows under hypothesis $\mathcal{H}_1$ which are synthetic gradient flow and the sampled version of the user transition flows, respectively.

For the Chicago Road network, we consider a synthetic curl flow under hypothesis $\mathcal{H}_0$ and two different flows under the hypothesis $\mathcal{H}_1$ which is a synthetic low-pass gradient flow with ith element of the embedding $\hat{\mathbf{x}}_{G,i} \sim \mathcal{N}\left(\exp(-i/5)^\top, 0.01\right)$ and the sampled version of the curl flows. Here, under $\mathcal{H}_1 - synthetic$, the prior is that the signal is low-pass.

Since these detectors are energy-based, we keep the flows under $\mathcal{H}_1 - synthetic$ with the same energy as those under $\mathcal{H}_0$ for a fair setting. The sampled version of flows under the $\mathcal{H}_1 - sampled$ simulates a setup when not all the flows are observed. And we want to detect from these partial measurements weather the observed flow is anomalous or not. The sampling rate is 50% for all experiments and the missing values in the sampling are filled with zero.

As for the baseline, it is intuitive to transform the edge flows into node signals in corresponding line-graph [17] and exploit graph-based detectors. Therefore, we build line graphs and consider the blind-SMSD (B-SMSD) [20] as a baseline. For the bandwidth of the B-SMSD we pick 20% of the line-graph eigenvectors corresponding to the eigenvalues with smaller magnitudes to form the band and compare the out-of-band signal-to-noise ratio with the threshold γ following [20].

C. Performance Comparision

TABLE II shows the experimental results. The area under curves (AUCs) of the receiver operating curves ($\mathrm{P_D}$ vs $\mathrm{P_{FA}}$) of the SHD for the theoretical and the experimental results concord in all the cases, thereby affirming the validity of the theoretical framework. The baseline graph detector B-SMSD

[1] $\|\hat{\mathbf{f}}_{\overline{\Delta}}\|_2^2$ is the signal embedding energy and $\|\hat{\mathbf{f}}_{\overline{\Delta}} - \hat{\mathbf{x}}_{\overline{\Delta}}^{*}\|_2^2$ can be regarded as the corrected term which is the energy difference between MLE $\hat{\mathbf{f}}_{\overline{\Delta}}$ and the constrained MLE $\hat{\mathbf{x}}_{\overline{\Delta}}^{*}$.

TABLE I: Experimental setup. The sampling rate in $\mathcal{H}_1 - sampled$ is 50% and the missing values in the sampling are filled with zero. The ith diagonal element of the matrix $\mathbf{R}$ is $\exp(i/20)$ and $\mu = 1$ for all datasets.

Dataset	$\mathcal{H}_0$	$\mathcal{H}_1 - synthetic$	$\mathcal{H}_1 - sampled$	**SNR**	$\mathbf{U}_{\overline{\Delta}}$
Forex [6]	Gradient flow	$\mathbf{x} = \mathbf{B}_2\mathbf{t}$, $\mathbf{t} \sim \mathcal{N}\left(\mathbf{0}, 0.3 \times 10^{-1}\mathbf{I}\right)$	Sampled flow in $\mathcal{H}_0$	-12dB	$[\mathbf{U}_{\mathrm{C}}\ \mathbf{U}_{\mathrm{H}}]$
Lastfm [6]	Non-gradient flow	$\mathbf{x} = \mathbf{B}_1^{\top}\mathbf{v}$, $\mathbf{v} \sim \mathcal{N}\left(\mathbf{0}, 5.0 \times 10^{1}\mathbf{I}\right)$	Sampled flow in $\mathcal{H}_0$	-12dB	$\mathbf{U}_{\mathrm{G}}$
Chicago [16]	$\mathbf{x} = \mathbf{B}_2\mathbf{t}$, $\mathbf{t} \sim \mathcal{N}\left(\mathbf{0}, 2.3 \times 10^{-2}\mathbf{I}\right)$	$\hat{\mathbf{x}}_{G,i} \sim \mathcal{N}\left(\exp(-i/5)^{\top}, 0.01\right)$	Sampled flow in $\mathcal{H}_0$	-15dB	$\mathbf{U}_{\mathrm{G}}$

TABLE II: Area under the curves (AUC). SHD-Th. and SHD-Exp. represent the theoretical and empirical results. Values in bold represent the model that achieved the best results in the experiment (theoretical results are not included).

Dataset		SHD-Th.	SHD-Exp.	CSHD	B-SMSD [20]
Forex	Syn.	0.71	**0.71**	0.56	0.45
	Samp.	0.55	**0.55**	0.51	0.37
Lastfm	Syn.	0.99	**0.99**	0.50	0.39
	Samp.	0.59	**0.59**	0.51	0.22
Chicago	Syn.	0.75	0.75	**0.85**	0.52
	Samp.	0.53	**0.53**	0.51	0.41

fails to detect the anomalies since it fails to discriminate and merges the different edge flow components, ultimately, confirming that the line-graph is not a suitable choice for this task.

The CSHD shows its effectiveness of detecting the anomalies that existed in the simplicial signals. However, the CSHD should be considered only when the prior information is relevant for the task at hand. In the Forex and the Lastfm datasets, the embeddings are not low-pass. Thus, instead of improving the performance, the enforced low-pass prior reduces the CSHD's performance. When the prior information is precise as in the Chicago synthetic case, the AUC of the CSHD is the best. For the Chicago sampled case, the low-pass prior is not holding anymore. Thus, CSHD performs worse than SHD. The results validate the idea that incorporating prior knowledge enhances the accuracy of estimating $\hat{\mathbf{x}}_{\overline{\Delta}}$, thereby contributing to the improved performance.

V. Conclusions

This paper studied the problem of detecting different simplicial components from noisy observations where Hodge embeddings satisfy particular properties of interest such as divergence-free or curl-free. We took the edge flows as an example to discuss various simplicial components induced by the Hodge decomposition and their properties. To detect whether an edge flows is in some specific Hodge subspaces or not, we proposed two hypotheses and exploit the simplicial embeddings of these flows w.r.t. the underlying topology. Based on the orthogonality between different Hodge subspaces, we developed the simplicial Hodge detector, which compares the energy of the subspaces' projections with the threshold to make a decision. Then, we considered prior information to enhance the detector performance and proposed the constrained simplicial Hodge detector. The numerical experiments verified the effectiveness of these two detectors and indicated that they are more suitable for simplicial edge flow detection compared with graph detectors.

References

[1] G. Bianconi, *Higher-order networks*. Cambridge University Press, 2021.

[2] C. Bick, E. Gross, H. A. Harrington, and M. T. Schaub, "What are higher-order networks?" *SIAM Review*, vol. 65, no. 3, pp. 686–731, 2023.

[3] F. Battiston, G. Cencetti, I. Iacopini, V. Latora, M. Lucas, A. Patania, J.-G. Young, and G. Petri, "Networks beyond pairwise interactions: Structure and dynamics," *Physics Reports*, vol. 874, pp. 1–92, 2020.

[4] V. Salnikov, D. Cassese, and R. Lambiotte, "Simplicial complexes and complex systems," *European Journal of Physics*, vol. 40, no. 1, p. 014001, 2018.

[5] C. Berge, *Hypergraphs: combinatorics of finite sets*. Elsevier, 1984, vol. 45.

[6] J. Jia, M. T. Schaub, S. Segarra, and A. R. Benson, "Graph-based semi-supervised & active learning for edge flows," in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp. 761–771.

[7] R. Money, J. Krishnan, B. Beferull-Lozano, and E. Isufi, "Online edge flow imputation on networks," *IEEE Signal Processing Letters*, vol. 30, pp. 115–119, 2022.

[8] M. Yang, E. Isufi, M. T. Schaub, and G. Leus, "Simplicial convolutional filters," *IEEE Transactions on Signal Processing*, vol. 70, pp. 4633–4648, 2022.

[9] M. Yang and E. Isufi, "Simplicial trend filtering," in *2022 56th Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2022, pp. 930–934.

[10] M. Yang, E. Isufi, and G. Leus, "Simplicial convolutional neural networks," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 8847–8851.

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

[12] J. Krishnan, R. Money, B. Beferull-Lozano, and E. Isufi, "Simplicial vector autoregressive model for streaming edge flows," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5.

[13] L.-H. Lim, "Hodge laplacians on graphs," *Siam Review*, vol. 62, no. 3, pp. 685–715, 2020.

[14] 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, p. 108149, 2021.

[15] E. Isufi and M. Yang, "Convolutional filtering in simplicial complexes," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 5578–5582.

[16] C. Liu, G. Leus, and E. Isufi, "Unrolling of simplicial elasticnet for edge flow signal reconstruction," *IEEE Open Journal of Signal Processing*, 2023.

[17] M. T. Schaub and S. Segarra, "Flow smoothing and denoising: Graph signal processing in the edge-space," in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2018, pp. 735–739.

[18] L. L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Transactions on signal processing*, vol. 42, no. 8, pp. 2146–2157, 1994.

[19] C. Hu, J. Sepulcre, K. A. Johnson, G. E. Fakhri, Y. M. Lu, and Q. Li, "Matched signal detection on graphs: Theory and application to brain imaging data classification," *NeuroImage*, vol. 125, pp. 587–600, 2016.

[20] E. Isufi, A. S. Mahabir, and G. Leus, "Blind graph topology change detection," *IEEE Signal Processing Letters*, vol. 25, no. 5, pp. 655–659, 2018.

[21] J. M. Górriz, J. Ramírez, C. G. Puntonet, and J. C. Segura, "Generalized lrt-based voice activity detector," *IEEE Signal Processing Letters*, vol. 13, no. 10, pp. 636–639, 2006.