Topological Adaptive Learning over Cell Complexes

Lorenzo Marinucci[1], Claudio Battiloro[2], Paolo Di Lorenzo[1]
[1]Department of Information Engineering, Electronics, and Telecommunications, Sapienza University, Rome, Italy.
[2]Harvard T.H. Chan School of Public Health, Harvard University, Boston, U.S.A.
Emails: lorenzo.marinucci96@gmail.com, cbattiloro@hsph.harvard.edu, paolo.dilorenzo@uniroma1.it

***Abstract*—This paper introduces a novel approach to adaptive learning from streaming flow signals defined over cell complexes, utilizing a topology-based least mean squares (LMS) strategy. By harnessing the principles of Hodge theory, we develop a topological LMS algorithm that efficiently gathers and integrates flow data across various edges and their neighboring cells at multiple levels. Through comprehensive theoretical examination, we elucidate the algorithm's stochastic behavior, outlining conditions that ensure stability in terms of mean and mean-square error. Furthermore, we derive explicit formulas for assessing the mean-square performance, highlighting how it is influenced by the underlying topological structure, sampling techniques, and data characteristics. Our empirical evaluations, using both synthetic and real-world network traffic datasets, validate our theoretical finding and demonstrate the superiority of our topological approach over traditional graph-based adaptive learning methods that overlook higher-order topological elements.**

I. Introduction

Over the past few years, a plethora of processing and learning techniques for signals on irregular, and not always metric, spaces have emerged. Graph Signal Processing (GSP) is a notable example, offering a suite of methods for the analysis and processing of signals on graph vertices. This includes the creation of various graph operators, which have paved the way for an array of graph filters and Fourier transforms [1]. The unique aspect of these tools is their reliance on graph connectivity, embedded within the structure of the chosen graph shift operators. While graph-based methods have become a cornerstone in data analysis, their intrinsic limitation is their focus on capturing only pairwise relationships between data points. This constraint makes them less effective in contexts where the dynamics involve more complex interactions that cannot be neatly encapsulated by simple pairwise connections [2]. Notably, this shortfall becomes apparent in fields such as biological networks [2], where intricate multi-way interactions among various entities like genes and proteins are the norm, or in neuroscience, where the simultaneous activation of neuron groups in brain networks presents a level of complexity beyond pairwise interactions [3]. These instances necessitate processing tools that extend beyond the capabilities of GSP, leading to the rise of Topological Signal Processing (TSP) [3]. The foundation of TSP was laid by pioneering research that highlighted the advantages of processing signals on topological spaces, such as simplicial or cell complexes, known as topological signals [3], [4]. The main advantage comes from a solid algebraic framework associated with such topological spaces. Subsequent research in TSP has made significant strides, including the development of FIR filters for these complex signals [5], [6], and the creation of a generalized Laplacian for mapping simplicial complexes onto traditional graphs [7]. Moreover, TSP has inspired the creation of advanced neural network architectures designed to learn from data defined in topological domains [8], [9].

A fundamental problem in signal processing is *adaptive learning*, whose aim is to infer and track the structure of an unknown system from streaming and noisy data observed over time [10]. Such powerful techniques have been largely studied in networked scenarios, with the aim of designing adaptive networks capable of distributed optimization and learning capabilities, see, e.g., [11] and references therein. More recently, adaptive learning methods have also been extended to process and learn from streaming signals defined over graphs [12]–[16]. In particular, the works in [12]–[14] recast classical adaptive estimation strategies in the graph signal processing framework, namely the LMS, the recursive least squares (RLS), and the Kalman filter, proposing effective node sampling techniques to enable adaptive reconstruction capability and guaranteed steady-state performance. Instead, the works in [15], [16] focused on distributed identification of graph filters from streaming nodal data, hinging on diffusion adaptation techniques. The aforementioned adaptive techniques apply to graph-based data, and are not suitable to exploit the richer structure offered by more general topological spaces such as, e.g., simplicial or cell complexes. To the best of our knowledge, only a few papers deal with online estimation and learning from streaming topological signals. Notable examples include simplicial vector autoregressive models for edge flows [17], and edge flow prediction over expanding simplicial complexes [18]. However, to current date, the field of *adaptive topological signal processing* is still a largely uncharted territory that is worth to be investigated.

In this work, we aim at providing a fundamental contribution in the field of adaptive TSP, revisiting the classical LMS algorithm to process and learn from streaming signals defined over topological domains. Specifically, hinging on formal arguments from algebraic topology [19], we generalize the work in [15] for graphs to regular cell complexes, thus introducing a novel class of topological LMS algorithms that are computational efficient and take into account the structure of the cell complex during processing. A detailed theoretical analysis provides the stochastic behavior of the proposed

This work was supported by the European Union under the Italian National Recovery and Resilience Plan (NRRP) of NextGenerationEU, partnership on "Telecommunications of the Future" (PE00000001 - program "RESTART").

algorithm, providing conditions for stability in the mean and mean-square error sense, and deriving closed-form expressions for the mean-square performance that depend on topological features, sampling, and data statistics. Finally, we illustrate the advantages of the proposed topological adaptive learning strategy through numerical experiments.

II. Background

This section introduces the essentials of topological signal processing over regular cell complexes, simplifying content while maintaining mathematical integrity.

Regular Cell Complex. A *regular cell complex* is a topological space $\mathcal{X}$ with a partition $\{\mathcal{X}_\sigma\}_{\sigma \in \mathcal{P}_\mathcal{X}}$ of subspaces $\mathcal{X}_\sigma$ of $\mathcal{X}$ (*cells*), where $\mathcal{P}_\mathcal{X}$ is the indexing set of $\mathcal{X}$, such that [19]:

1) For each $c \in \mathcal{X}$, every sufficient small neighborhood of c intersects finitely many $\mathcal{X}_\sigma$;
2) For all τ,σ we have that $\mathcal{X}_\tau \cap \overline{\mathcal{X}}_\sigma \neq \varnothing$ iff $\mathcal{X}_\tau \subseteq \overline{\mathcal{X}}_\sigma$, where $\overline{\mathcal{X}}_\sigma$ is the closure of the cell;
3) Every $\mathcal{X}_\sigma$ is homeomorphic to $\mathbb{R}^k$ for some k;
4) For every $\sigma \in \mathcal{P}_\mathcal{X}$ there is a homeomorphism ϕ of a closed ball in $\mathbb{R}^k$ to $\overline{\mathcal{X}}_\sigma$ such that the restriction of ϕ to the interior of the ball is a homeomorphism onto $\mathcal{X}_\sigma$.

Condition 2 implies that $\mathcal{P}_\mathcal{X}$ has a poset structure, given by $\tau \leq \sigma$ iff $\mathcal{X}_\tau \subseteq \overline{\mathcal{X}_\sigma}$, and we say that τ *bounds* σ. Condition 4 implies that all of the topological information about $\mathcal{X}$ is encoded in the poset structure of $\mathcal{P}_\mathcal{X}$. For this reason, we will indicate the cell $\mathcal{X}_\sigma$ with its corresponding poset element σ. The dimension dim(σ) of a cell σ is k, we call it a $k-$cell and denote it with σ^k. Regular cell complexes can be described via a consistent boundary relation.

Boundary Relation. We say that σ is on the boundary of τ, denoted as $\sigma \prec \tau$, iff $\dim(\sigma) \leq \dim(\tau)$ and there is no cell δ such that $\sigma \leq \delta \leq \tau$.

Therefore, the boundary of a cell σ^k of dimension k is the set of all cells of dimension less than k bounding σ^k. The order of a cell complex is the largest dimension of any of its cells and we denote an order K regular cell complex with $\mathcal{X}^K$. A graph is a particular case of a cell complex of dimension 1, containing only cells of dimension 0 (nodes) and 1 (edges). A graph with 2-dimensionals cells being some of its induced cycles, referred to as *polygons*, is a cell complex of order 2. We denote the set of k-cells in $\mathcal{X}^K$ as $\mathcal{D}_k := \{\sigma_i^k : \sigma_i^k \in \mathcal{X}_K\}$.

Topological Signals. A k-topological signal $\mathbf{x}_k$ over a regular cell complex $\mathcal{X}^K$ is defined as a collection of mappings from the set of all k-cells contained in the complex to real numbers:

$$\mathbf{x}_k = [x_k(\sigma_1^k), \ldots, x_k(\sigma_i^k), \ldots, x_k(\sigma_{Q_k}^k)] \in \mathbb{R}^{Q_k}, \quad (1)$$

where $x_k : \mathcal{D}_k \to \mathbb{R}$ and , $Q_k = |\mathcal{D}_k|$.

Assume that a reference orientation of the complex is given, detailed explanations can be found in [6], [20]. There are two ways in which two cells can be considered to be adjacent: lower and upper adjacent. Two $k-$cells are lower adjacent if they share a common face of dimension $k-1$ and upper adjacent if both are faces of a cell of dimension $k+1$. The structure of an oriented regular cell complex of order K is then fully captured by the set of its incidence (or boundary) matrices $\mathbf{B}_k \in \mathbb{R}^{Q_{k-1} \times Q_k}$, $k = 1, \ldots, K$, with entries $B_k(i,j) = 0$ if σ_i^{k-1} is not a face of σ_j^k, and $B_k(i,j) = 1$ (or -1), if σ_i^{k-1} is a face of σ_j^k and its orientation is coherent (or not) with the orientation of σ_j^k.

Hodge Laplacians. From the incidence information, we build the Hodge Laplacian matrices as [3]:

$$\mathbf{L}_0 = \mathbf{B}_1 \mathbf{B}_1^T, \quad (2)$$

$$\mathbf{L}_k = \underbrace{\mathbf{B}_k^T \mathbf{B}_k}_{\mathbf{L}_{k,d}} + \underbrace{\mathbf{B}_{k+1} \mathbf{B}_{k+1}^T}_{\mathbf{L}_{k,u}}, \quad k = 1, \ldots, K-1, \quad (3)$$

$$\mathbf{L}_K = \mathbf{B}_K^T \mathbf{B}_K. \quad (4)$$

Laplacian matrices of intermediate dimensions, i.e., $k = 1, \ldots, K-1$, contain two terms. The first term $\mathbf{L}_{k,d}$ is the lower Laplacian, and it encodes the lower connectivity among k-dimensional cells; the second term $\mathbf{L}_{k,u}$ is the upper Laplacian and it encodes the upper connectivity among k-dimensional cells. For example, two edges are upper adjacent if they are faces of a common polygon, or lower adjacent if they share a common vertex.

Cell Complex FIR Filters. Given the k-th Hodge Laplacian $\mathbf{L}_k$ as in (3), a cell complex FIR filter acting on $k-$topological signals is defined as a polynomial of the Laplacian as [5]:

$$\mathbf{H}_k = \sum_{m=0}^{M} h_{m,u} \big(\mathbf{L}_{k,u}\big)^m + \sum_{m=1}^{M} h_{m,d} \big(\mathbf{L}_{k,d}\big)^m, \quad (5)$$

where M is a positive integer and $\{h_{m,u}\}_{m=0}^M \in \mathbb{R}^{M+1}$, $\{h_{m,d}\}_{m=1}^M \in \mathbb{R}^M$. In the sequel, we focus on the case $k=1$, neglecting the subscript k for the sake of clarity (e.g., we indicate $\mathbf{L}_1$ with $\mathbf{L}$, $\mathbf{x}_1$ with $\mathbf{x}$, and Q_1 with Q).

III. Topological Least Mean Squares Adaptive Filtering

Consider a cell complex consisting of a set of Q edges, and a set of P polygons. The topological information of the cell complex can be encoded into the structure of the upper Laplacian $\mathbf{L}_u$ and the lower Laplacian $\mathbf{L}_d$, as explained in Sec. II. In this paper, we are interested in the adaptive processing of flow signals (partially) observed over the edges of the cell complex. Let $\mathbf{x}(n)$ be stationary edge flow signals, processed over time by the linear shift-invariant graph filters in (5) according to the following model:

$$\boldsymbol{y}(n) = \mathbf{D}(n) \Bigg[\sum_{m=0}^{M} h_{m,u}^o \big(\mathbf{L}_u\big)^m \mathbf{x}(n-m) + \sum_{m=1}^{M} h_{m,d}^o \big(\mathbf{L}_d\big)^m \mathbf{x}(n-m) + \boldsymbol{v}(n) \Bigg], \quad (6)$$

with $n \geq M$, where $\boldsymbol{v}(n) = [v_1(n), \ldots, v_Q(n)]^T$ denotes an $Q \times 1$ i.i.d. zero-mean measurement noise, independent of any other signal and with covariance matrix $\boldsymbol{R}_v$; $\boldsymbol{h}^o = [h_{u,0}^o, h_{1,u}^o, \ldots, h_{M,u}^o, h_{1,d}^o, \ldots, h_{M,d}^o]^T \in \mathbb{R}^{2M+1}$ is the vector of topological filter coefficients to be estimated; and $\mathbf{D}(n) = \mathrm{diag}(d_1(n), \ldots, d_Q(n)) \in \mathbb{R}^{Q \times Q}$ is a sampling operator such that $d_i(n) = 1$ if edge i is sampled at time n, and zero otherwise. It should be noted that, although in (5) variable m is just an index for the powers in the polynomial, it defines also time-lags in (6). The model in (6) generalizes the one proposed in [15], considering noisy and partial observations of two graph shift operations that combine streaming edge data over upper and lower neighborhoods, respectively. In the sequel, we make an assumption on the sampling process.

Assumption 1 (Independent sampling): The random variables extracted from the sampling process $d_i(l)$ are temporally and spatially independent, for all i, and $l \leq n$.

We also assume that the signal $\mathbf{x}(n)$ is a zero-mean wide-sense stationary process, i.e., $\mathbb{E}\{\mathbf{x}(n)\} = \mathbf{0}$ for all n, and its autocorrelation sequence $\mathbf{R}_x(m)$=$\mathbb{E}\{\mathbf{x}(n)\mathbf{x}^T(n-m)\}$ is a function of the time lag m only. Our goal is to estimate the filter coefficients $\mathbf{h}^o$ in (6). To this aim, we consider the mean-square-error criterion:

$$\min_{\mathbf{h}} \ \mathbb{E}\{\|\boldsymbol{y}(n) - \mathbf{D}(n)\mathbf{X}(n)\mathbf{h}\|^2\} \tag{7}$$

where $\mathbf{X}(n)$ is an $Q \times 2M+1$ matrix given by:

$$\begin{aligned}\mathbf{X}(n) = \big[&\mathbf{x}(n), \mathbf{L}_u\mathbf{x}(n-1), \ldots, \mathbf{L}_u^M\mathbf{x}(n-M),\\ &\mathbf{L}_d\mathbf{x}(n-1), \ldots, \mathbf{L}_d^M\mathbf{x}(n-M)\big],\end{aligned} \tag{8}$$

which collects shifted versions of the edge flows $\mathbf{x}(n-m)$, $m = 1, \ldots, M$, over the upper and lower neighborhoods. By setting the gradient vector of (7) to zero, the optimal parameter vector $\mathbf{h}^o$ can be found by solving:

$$\mathbf{R}_X\mathbf{h}^o = \mathbf{r}_{Xy}, \tag{9}$$

where the $2M+1 \times 2M+1$ matrix $\mathbf{R}_X$ and the $2M+1 \times 1$ vector $\mathbf{r}_{Xy}$ are given by:

$$\begin{aligned}\mathbf{R}_X &= \mathbb{E}\{\mathbf{X}(n)^T\mathbf{D}(n)\mathbf{X}(n)\}\\ \mathbf{r}_{Xy} &= \mathbb{E}\{\mathbf{X}(n)^T\mathbf{D}(n)\boldsymbol{y}(n)\}.\end{aligned}$$

The matrix $\mathbf{R}_X$ and the vector $\mathbf{r}_{Xy}$ do not depend on n, due to the stationarity of $\mathbf{x}(n)$ and Assumption 1. Let $\mathbf{p} = (p_1, \ldots, p_Q)^T \in \mathbb{R}^Q$ represent the sampling probability vector, with $p_i = \mathbb{E}\{d_i(n)\}$, for $i = 1, \ldots, Q$, denoting the probability that node i is sampled at time n. Let $\mathbf{P} = \mathrm{diag}(\boldsymbol{p})$. The matrix $\mathbf{R}_X$ is a 2×2 block matrix:

$$\mathbf{R}_X = \begin{bmatrix}\mathbf{R}_{u,X} & \mathbf{R}_{ud,X}\\ \mathbf{R}_{ud,X} & \mathbf{R}_{d,X}\end{bmatrix} \tag{10}$$

where the (m, l) component of matrix $\mathbf{R}_{u,X}$ is given by

$$\begin{aligned}[\mathbf{R}_{u,X}]_{m,l} &= \mathbb{E}\left\{\mathbf{x}^T(n-m)(\mathbf{L}_u^m)^T\mathbf{D}(n)\mathbf{L}_u^l\mathbf{x}(n-l)\right\}\\ &= \mathrm{Tr}\left((\mathbf{L}_u^m)^T\mathbf{P}\mathbf{L}_u^l\mathbf{R}_x(m-l)\right).\end{aligned} \tag{11}$$

Similarly, we have:

$$\begin{aligned}[\mathbf{R}_{d,X}]_{m,l} &= \mathrm{Tr}\left((\mathbf{L}_d^m)^T\mathbf{P}\mathbf{L}_d^l\mathbf{R}_x(m-l)\right),\\ [\mathbf{R}_{ud,X}]_{m,l} &= \mathrm{Tr}\left((\mathbf{L}_u^m)^T\mathbf{P}\mathbf{L}_d^l\mathbf{R}_x(m-l)\right).\end{aligned} \tag{12}$$

Finally, $\mathbf{r}_{Xy} = [\mathbf{r}_{Xy}^u, \mathbf{r}_{Xy}^d]$ is a block vector with components

$$\begin{aligned}\mathbf{r}_{Xy}^u &= \mathrm{Tr}\left((\mathbf{L}_u^m)^T\mathbf{P}\mathbf{R}_{xy}(m)\right),\\ \mathbf{r}_{Xy}^d &= \mathrm{Tr}\left((\mathbf{L}_d^m)^T\mathbf{P}\mathbf{R}_{xy}(m)\right),\end{aligned} \tag{13}$$

Algorithm 1 Topo-LMS: Topological Least Mean Squares on Cell Complexes

```
Initialize h(0) randomly
Given a sufficiently small step-size μ > 0
for each time n ≥ 0 do
    Choose a sampling operator D(n);
    Observe y(n), and build X(n) as in (8);
    Perform parameter adaptation as in (16).
```

where $\mathbf{R}_{xy}(m) = \mathbb{E}\{\boldsymbol{y}(n)\mathbf{x}^T(n-m)\}$ denotes the cross correlation function, which is independent of the time index n. The system of equations (9) admits a unique solution if the square matrix $\mathbf{R}_X$ is full rank, i.e.,

$$\lambda_{\min}(\mathbf{R}_X) > 0,$$

where $\lambda_{\min}(\mathbf{Y})$ is the minimum eigenvalue of matrix $\mathbf{Y}$. From (11)-(12), it is clear that this condition depends strongly on the selection of the sampling probability matrix $\mathbf{P}$. In this case, $\mathbf{h}^o$ can be determined uniquely by either solving (9), or proceeding iteratively using gradient descent:

$$\mathbf{h}(n+1) = \mathbf{h}(n) + \mu\big(\mathbf{r}_{Xy} - \mathbf{R}_X\mathbf{h}(n)\big) \tag{14}$$

where $\mu > 0$ is a (sufficiently small) step-size. Since the second order moments are rarely available beforehand, or possibly time-varying, it is necessary to approximate them via instantaneous approximations, with different constructions leading to different adaptive algorithms. One of the simplest choices is to use the instantaneous approximations

$$\mathbf{R}_X \approx \mathbf{X}^T(n)\mathbf{D}(n)\mathbf{X}(n), \quad \mathbf{r}_{Xy} \approx \mathbf{X}^T(n)\mathbf{D}(n)\boldsymbol{y}(n). \tag{15}$$

This leads to the following adaptive filter:

$$\mathbf{h}(n+1) = \mathbf{h}(n) + \mu\mathbf{X}^T(n)\mathbf{D}(n)\big(\boldsymbol{y}(n) - \mathbf{X}(n)\mathbf{h}(n)\big) \tag{16}$$

that is referred to as the topological LMS algorithm (Topo-LMS, cf. Algorithm 1), where streaming flow signals $\{\boldsymbol{y}(n), \mathbf{x}(n)\}$ are processed taking into account the topological information coming from the cell complex domain.

IV. Mean Square Analysis

In this section, we briefly present the performance in the mean and mean-square-error sense of the Topo-LMS algorithm. Let $\widetilde{\mathbf{h}}(n)$ denote the error vector defined as:

$$\widetilde{\mathbf{h}}(n) = \mathbf{h}^o - \mathbf{h}(n). \tag{17}$$

Using the data model (6) and following the same line of reasoning as in [10], the error vector evolves according to:

$$\widetilde{\mathbf{h}}(n+1) = \mathbf{B}(n)\widetilde{\mathbf{h}}(n) - \mu\boldsymbol{g}(n) \tag{18}$$

where

$$\mathbf{B}(n) = \mathbf{I} - \mu\mathbf{X}(n)^T\mathbf{D}(n)\mathbf{X}(n) \tag{19}$$

$$\boldsymbol{g}(n) = \mathbf{X}(n)^T\mathbf{D}(n)\boldsymbol{v}(n) \tag{20}$$

In the sequel, to simplify the analysis, we introduce the following assumption on the regressors $\{\mathbf{D}(n)\mathbf{X}(n)\}_n$.

Assumption 2: The regressors $\mathbf{D}(n)\mathbf{X}(n)$ arise from a zero-mean random process that is temporally white with positive-definite covariance $\mathbf{R}_X$.

This assumption means that $\mathbf{D}(n)\mathbf{X}(n)$ is independent of $\mathbf{D}(j)\mathbf{X}(j)$ for all $j < n$. Despite this assumption might not be true in general, it is commonly used when analyzing adaptive constructions since it allows to simplify the derivations without constraining the conclusions [10].

Taking expectation of both sides of (18), and using Assumptions 1 and 2, and the fact that $\mathbb{E}\{\boldsymbol{g}(n)\} = \mathbf{0}$, we obtain:

$$\mathbb{E}\{\widetilde{\mathbf{h}}(n+1)\} = \mathbf{B}\,\mathbb{E}\{\widetilde{\mathbf{h}}(n)\},$$

where

$$\mathbf{B} = \mathbf{I} - \mu\mathbf{R}_X \tag{21}$$

with $\mathbf{R}_X$ having the structure in (10)-(12). Thus, the estimates $\{\mathbf{h}(n)\}$ generated by algorithm (16) converge in the mean to the optimal solution $\mathbf{h}^o$ if $\mathbf{B}$ is stable, namely, the spectral radius $\rho(\mathbf{B}) < 1$. From (21), the stability of $\mathbf{B}$ is ensured if the step-size μ satisfies:

$$0 < \mu < \frac{2}{\lambda_{\max}(\mathbf{R}_X)}. \tag{22}$$

The mean-square-error behavior of algorithm (16) can be characterized by studying the evolution of $\mathbb{E}\{\|\widetilde{\mathbf{h}}(n)\|^2_{\boldsymbol{\Sigma}}\}$ for any positive semi-definite matrix $\boldsymbol{\Sigma}$ that we are free to choose. From recursion (16) and using the independence Assumptions 1 and 2, the weighted variance evolves according to

$$\mathbb{E}\{\|\widetilde{\mathbf{h}}(n+1)\|^2_{\boldsymbol{\Sigma}}\} = \mathbb{E}\{\|\widetilde{\mathbf{h}}(n)\|^2_{\boldsymbol{\Sigma}'}\} + \mu^2\mathrm{Tr}(\boldsymbol{\Sigma}\mathbf{G}) \tag{23}$$

with

$$\boldsymbol{\Sigma}' = \mathbb{E}\{\mathbf{B}^T(n)\boldsymbol{\Sigma}\mathbf{B}(n)\}, \tag{24}$$

$$\mathbf{G} = \mathbb{E}\{\boldsymbol{g}(n)\boldsymbol{g}^T(n)\} = \sigma_v^2\mathbf{R}_X, \tag{25}$$

where we used the spatially independence assumption of the zero-mean noise $\boldsymbol{v}(n)$. Let $\boldsymbol{\sigma} = \mathrm{vec}(\boldsymbol{\Sigma})$ denote the vector representation of $\boldsymbol{\Sigma}$ obtained by stacking the columns entries of $\boldsymbol{\Sigma}$ on top of each other. Let $\boldsymbol{\sigma}' = \mathrm{vec}(\boldsymbol{\Sigma}')$. The vector $\boldsymbol{\sigma}'$ can be related to $\boldsymbol{\sigma}$ according to $\boldsymbol{\sigma}' = \mathbf{F}\boldsymbol{\sigma}$ where $\mathbf{F} = \mathbb{E}\{\mathbf{B}^T(n) \otimes \mathbf{B}^T(n)\}$, with $\otimes$ denoting the Kronecker product. The evaluation of $\mathbf{F}$ requires knowledge of the fourth-order moments of the flow signals, which are typically not available. A common alternative is to use the approximation $\mathbf{F} \approx \mathbf{B}^T \otimes \mathbf{B}^T$ for sufficiently small step-sizes [10], [11]. Under this approximation, the stability of $\mathbf{F}$ is ensured if $\rho(\mathbf{B}) < 1$, i.e., if the step-sizes are chosen according to (22). Then, the variance relation (23) can be recast as:

$$\mathbb{E}\{\|\widetilde{\mathbf{h}}(n+1)\|^2_{\boldsymbol{\sigma}}\} = \mathbb{E}\{\|\widetilde{\mathbf{h}}(n)\|^2_{\mathbf{F}\boldsymbol{\sigma}}\} + \mu^2\mathrm{vec}\,(\mathbf{G})^T\,\boldsymbol{\sigma} \tag{26}$$

where the notation $\|\mathbf{x}\|^2_{\boldsymbol{\sigma}}$ is used to denote the same quantity $\|\mathbf{x}\|^2_{\boldsymbol{\Sigma}} = \mathbf{x}^T\boldsymbol{\Sigma}\mathbf{x}$. If $\mathbf{F}$ is stable, the variance recursion converges to a limit point [11]:

$$\lim_{n\to\infty} \mathbb{E}\{\|\widetilde{\mathbf{h}}(n)\|^2_{\boldsymbol{\sigma}}\} = \mu^2\mathrm{vec}\,(\mathbf{G})^T\,(\mathbf{I}-\mathbf{F})^{-1}\boldsymbol{\sigma}. \tag{27}$$

From (27), we can obtain the mean-square deviation (MSD) by simply choosing $\boldsymbol{\sigma} = \mathrm{vec}(\mathbf{I})$.

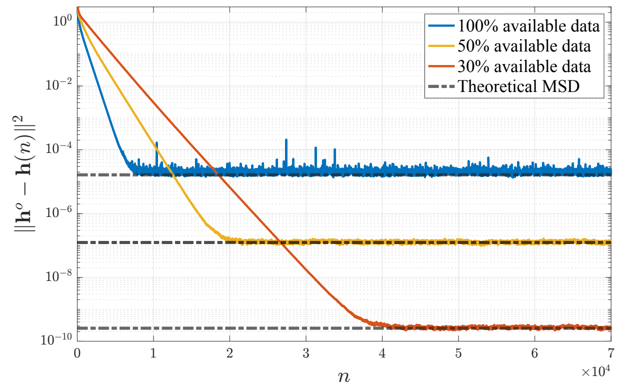

Fig. 1: MSD vs iteration index, for different percentages of observed edges, compared with theoretical expressions.

V. Numerical Results

We first assess the performance of the proposed LMS adaptive filtering on synthetic data. We build a simplicial complex, i.e. a regular cell complex, with the following iterative procedure. Starting from a randomly selected number of nodes, the structure of the simplicial complex is enriched at each step with randomly generated simplices of increasing dimension. The resulting complex $\mathcal{X}^2$ includes 11 nodes, 20 edges and 7 triangles [3], and we process edge flow signals ($k = 1$ from (1)) defined on it. We generate a dataset using the model in (6), with the i.i.d. zero-mean measurement noise being AWGN. The variance of the noise is different on each edge, and it is randomly sampled from $\{10^{-6}, 10^{-3}, 10^{-1}\}$. We run our Topo-LMS algorithm from Alg. 1 aiming to recover the true filter coefficients $\mathbf{h}^o$ we used to generate the dataset. In Fig. 1, we show the MSD for three different sampling percentages, averaging over 100 realizations, where X% of available data means that only the X% of diagonal entries of $\mathbf{P}$ are greater than zero, i.e. only the X% of edges has a non-zero probability of being sampled. We also report the corresponding theoretical value from (27), which is clearly attained in all the cases. The stepsize μ is set equal to 10^{-4}. We designed the experiment such that the lowest sampling percentage (30%) assigns non-zero sampling probabilities to the edges with the lowest noise variance. In this way, from Fig. 1, we can appreciate the interesting tradeoff between learning rate and MSD at steady state; in particular, the fact that, in the 30% case, only the cleanest signal components are sampled, allows Topo-LMS to converge to a better MSD but with a slower learning rate, due to the slower diffusion over the complex. Vice versa, in the 100% case, the signal is completely available but more noisy, leading to a faster convergence but a worse MSD. This experiment, beyond validating the proposed algorithm and its theoretical analysis, motivates the design of principled sampling strategies for topological signals.

We now consider the German National Research and Education Network operated by the German DFN-Verein (DFN), being the communications infrastructure for Germany's broader scientific community. We model the network as a cell complex of order 2, taking all the induced cycles of the network as polygons [21]. The resulting complex consists of 50 nodes, 89

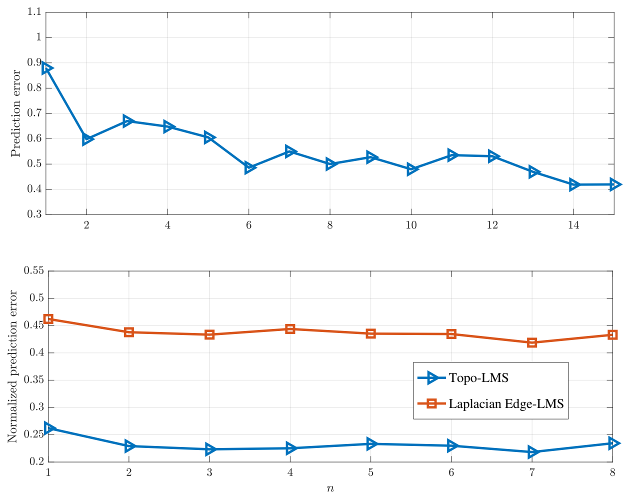

Fig. 2: (*Top*) Prediction error over training data. (*Bottom*) Performance comparison of different methods over test data.

edges, and 39 polygons. The data traffic is aggregated daily over February 2005, and the data measurements are expressed in Mbit/sec and collected on each link. We then have a time series made of 28 edge signals, the first 18 of whom we use for training, and the last 10 for testing. We don't sample the signals, i.e. $\mathbf{P} = \mathbf{I}$, and we employ the signal model in (1) in an autoregressive fashion. Thus, given the edge signal $\mathbf{y}(n+1)$ at day $n+1$ and the edge signals of previous days $\{\boldsymbol{y}(n-m), m = 0, \ldots, M-1\}$, Topo-LMS minimizes

$$\frac{1}{18-M}\sum_{n=M}^{17}\left\|\boldsymbol{y}(n+1) - \sum_{m=0}^{M} h_{m,u}\big(\mathbf{L}_u\big)^m \boldsymbol{y}(n-m) - \sum_{m=1}^{M} h_{m,d}\big(\mathbf{L}_d\big)^m \boldsymbol{y}(n-m)\right\|^2$$

w.r.t. $h_{m,u}$ and $h_{m,d}$, i.e. the autoregressive version of (7). For this experiment, we set the filter order M to 3, and the stepsize μ to 10^{-4}. We compare our Topo-LMS with the LMS algorithm resulting from using a signal model leveraging the Edge Laplacian [3], corresponding to (1) with only $\mathbf{L}_d$ in place of $\mathbf{L}$. The Edge Laplacian parametrization represents a fair comparison because it is the most naive generalization to cell complexes of the work [15] for graphs, not leveraging the Hodge theory. Then, in Fig. 2 (top), we show the convergence behavior of our method during training; whereas, in Fig. 2 (bottom) we illustrate the (normalized) daily prediction error obtained using the filter coefficients learnt during training over test traffic data from the 19th of February to the 28th of February (i.e., data we didn't use for training). As we can notice from Fig. 2 (bottom), our Topo-LMS is able to achieve a better prediction error w.r.t. to the graph-based approach, thanks to its ability to exploit higher-order topological information.

VI. Conclusions

In this work, we introduced an innovative adaptive learning technique for processing streaming flow signals on cell complexes, utilizing a topological least mean squares framework. By incorporating Hodge theory, our topological LMS algorithm efficiently processes flow data, maintaining stability in mean and mean-square errors under specific conditions. We detail the influence of topology, sampling strategies, and data characteristics on performance through explicit formulas. Empirical tests with synthetic and real-world data confirm our approach's effectiveness, showcasing its superiority over traditional graph-based methods.

References

[1] E. Isufi, F. Gama, D. I. Shuman, and S. Segarra, "Graph filters for signal processing and machine learning on graphs," *arXiv preprint arXiv:2211.08854*, 2022.
[2] 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.
[3] S. Barbarossa and S. Sardellitti, "Topological signal processing over simplicial complexes," *IEEE Trans. on Signal Processing*, vol. 68, pp. 2992–3007, 2020.
[4] 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.
[5] M. Yang, E. Isufi, M. T. Schaub, and G. Leus, "Simplicial convolutional filters," 2022.
[6] T. M. Roddenberry, M. T. Schaub, and M. Hajij, "Signal processing on cell complexes," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2022, pp. 8852–8856.
[7] F. Ji, G. Kahn, and W. P. Tay, "Signal processing on simplicial complexes with vertex signals," *IEEE Access*, vol. 10, pp. 41889–41901, 2022.
[8] 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.
[9] L. Giusti, C. Battiloro, P. Di Lorenzo, S. Sardellitti, and S. Barbarossa, "Simplicial attention neural networks," arXiv:2203.07485, 2022.
[10] A. H. Sayed, *Adaptive filters*, John Wiley & Sons, 2011.
[11] A. H. Sayed et al., "Adaptation, learning, and optimization over networks," *Foundations and Trends® in Machine Learning*, vol. 7, no. 4-5, pp. 311–801, 2014.
[12] P. Di Lorenzo, S. Barbarossa, P. Banelli, and S. Sardellitti, "Adaptive least mean squares estimation of graph signals," *IEEE Trans. on Signal and Information Proc. over Networks*, vol. 2, no. 4, pp. 555–568, 2016.
[13] P. Di Lorenzo, P. Banelli, E. Isufi, S. Barbarossa, and G. Leus, "Adaptive graph signal processing: Algorithms and optimal sampling strategies," *IEEE Trans. on Signal Processing*, vol. 66, no. 13, pp. 3584–3598, 2018.
[14] E. Isufi, P. Banelli, P. Di Lorenzo, and G. Leus, "Observing and tracking bandlimited graph processes from sampled measurements," *Signal Processing*, vol. 177, pp. 107749, 2020.
[15] R. Nassif, C. Richard, J. Chen, and A. H. Sayed, "Distributed diffusion adaptation over graph signals," in *Proc. of ICASSP*. IEEE, 2018, pp. 4129–4133.
[16] F. Hua, R. Nassif, C. Richard, H. Wang, and A. H. Sayed, "Online distributed learning over graphs with multitask graph-filter models," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 6, pp. 63–77, 2020.
[17] J. Krishnan, R. Money, B. Beferull-Lozano, and E. Isufi, "Simplicial vector autoregressive model for streaming edge flows," in *Proc. of IEEE ICASSP*, 2023, pp. 1–5.
[18] M. Yang, B. Das, and E. Isufi, "Online edge flow prediction over expanding simplicial complexes," in *Proc. of IEEE ICASSP*, 2023.
[19] L. J. Grady and J. R. Polimeni, *Discrete calculus: Applied analysis on graphs for computational science*, vol. 3, Springer, 2010.
[20] S. Sardellitti, S. Barbarossa, and L. Testa, "Topological signal processing over cell complexes," in *2021 55th Asilomar Conference on Signals, Systems, and Computers*, 2021, pp. 1558–1562.
[21] S. Orlowski, R. Wessäly, M. Pióro, and A. Tomaszewski, "Sndlib 1.0—survivable network design library," *Networks: An International Journal*, vol. 55, no. 3, pp. 276–286, 2010.