SIGNAL PROCESSING ON SIMPLICIAL COMPLEXES

Feng Ji    Giacomo Kahn    Wee Peng Tay

School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore

ABSTRACT

Theoretical development and applications of graph signal processing (GSP) have attracted much attention. In classical GSP, the underlying structures are restricted in terms of dimensionality. A graph is a combinatorial object that models binary relations, and it does not directly model complex \(n\)-ary relations. One possible high dimensional generalization of graphs are simplicial complexes. They are a step between the constrained case of graphs and the general case of hypergraphs. In this paper, we develop a signal processing framework on simplicial complexes, such that we recover the traditional GSP theory when restricted to signals on graphs. We demonstrate how to perform signal processing with the framework using numerical examples.

Index Terms— Graph signal processing, simplicial complex

1. INTRODUCTION

Many fields of research use data to create hypotheses and try to infer them. Due to this heterogeneous landscape, the data themselves come in diverse forms, from simple binary relations to relations of high arity. One way of incorporating topological properties in data analysis is to use graph signal processing (GSP) [1–8]. From signals recorded on networks such as sensor networks, GSP uses graph metrics that relay the topology of the graph to perform sampling, translating and filtering of the signals. Recently, GSP based graph convolution neural network also receives much attention [5, 9–11].

GSP, as useful a tool as it is, still has its limitations. The vast data landscape includes complex data, such as high dimensional manifolds, or point clouds possessing high dimensional geometric features (cf. Fig. 1). For example, 2D meshes can be used to approximate a surface and high dimensional simplicial complexes can be used to model discrete point clouds. Another example is in social networks such as Facebook, where an edge represents the friend relation but higher arity edges can represent family links or the inclusion in the same groups. This model also works for group conversations or other user groups in social networks. It is then necessary to go beyond graphs to fully capture these more complex interaction mechanisms. For these reasons, there is a need for a framework that permits signal processing on such high dimensional geometric objects.

Despite the fact that the subject is relatively new, a few attempts have been made. In [12, 13] the authors develop a signal processing framework using a differential operator on simplicial chain complexes. It considers signals associated with high dimensional simplices such as edges and faces, and not only vertices. The paper [14] proposes an approach on meet or join semi-lattices that uses lattice operators as the shift and [15] proposes a framework on hypergraphs using tensor decomposition.

In our paper, we propose another signal processing framework for signals on nodes of simplicial complexes. Our approach makes full use of the geometric structures and strictly generalizes traditional GSP by introducing generalized Laplacians. Signal processing tasks can thus be performed similar to traditional GSP. The rest of the paper is organized as follows. We recall fundamentals of simplicial complexes in Section 2. In Section 3, we introduce the generalized Laplacian. We focus on the special case of 2-complexes in Section 4. We present simulation results in Section 5 and conclude in Section 6. Proofs can be found in the extended report [16].

2. SIMPLICIAL COMPLEXES

In this section, we give a brief overview of the theory of simplicial complexes (see [17, 18] for more details).

The standard \(n\)-simplex (or dimension \(n\) simplex) \(\Delta_n\) is defined as \(\{x_0 + \ldots + x_n = 1 \mid (x_0, \ldots, x_n) \in \mathbb{R}^{n+1}_+\}\). Any topological space homeomorphic to the standard \(n\)-simplex is

---

This work was supported in part by the Singapore Ministry of Education Academic Research Fund Tier 2 grant MOE2018-T2-2-019. The computational work for this article was partially performed on resources of the National Supercomputing Centre, Singapore (https://www.nscc.sg).
called an \( n \)-simplex. In \( \Delta_n \), if we require \( k > 0 \) coordinates being 0, we obtain an \((n - k)\)-simplex, called a face.

A simplicial complex \( X \) (see Fig. 2 for an example) is a set of simplices such that any face from a simplex of \( X \) is also in \( X \) and the intersection for any two simplices \( \sigma_1, \sigma_2 \) of \( X \) is a face of both \( \sigma_1 \) and \( \sigma_2 \). A simplex of \( X \) is called maximal if it is not the face of any other simplices.

We shall primarily focus on finite simplicial complexes, i.e., a finite set of simplices. The dimension \( \dim X \) of \( X \) is the largest dimension of a simplex in \( X \). For each \( m \), the subset of \( m \)-simplices of \( X \) and their faces is called its \( m \)-skeleton, denoted by \( X^m \). The geometric realization of a simplicial complex \( X \) is the topological space obtained by gluing simplices with common faces. For example, a graph is the geometric realization of a 1-simplicial complex. For convenience, we shall not distinguish simplicial complex with its geometric realization when no confusion arises.

![Fig. 2](image)

(a) \( X \) is (the geometric realization) of a 3-complex with a maximal 3-simplex, 2 maximal 2-simplices and 3 maximal 1-simplices. (b) \( X^1 \) is a connected graph with 15 edges.

3. GENERALIZED LAPLACIAN

GSP relies heavily on the notion of a “shift operator”. A popular choice is the graph Laplacian. Now, we propose a generalization of this notion, extendable to any finite point cloud.

**Definition 1.** Let \( X \) be a finite simplicial complex. A generalized Laplacian consists of the following data:

(A) a weighted, undirected graph \( G_X = (V, E) \),

(B) a set function \( f : X^0 \to V \), and

(C) a linear transformation \( T : \mathbb{R}^{|X^0|} \to \mathbb{R}^{|V|} \),

such that the following holds:

(a) \( f \) is one-one,

(b) the \( f(v) \) component of \( T(x) \) is the same as the \( v \) component of \( x \) for each \( v \in X^0 \) and \( x \in \mathbb{R}^{|X^0|} \), and

(c) the sum of each row of \( T \) (written as a transformation matrix) is a constant.

Let \( L_{G_X} \) be the Laplacian of the weighted graph \( G_X \). The generalized Laplacian associated with the data \((G_X, f, T)\) is defined as

\[
L_{(G_X, f, T)} = T^* \circ L_{G_X} \circ T : \mathbb{R}^{|X^0|} \to \mathbb{R}^{|X^0|},
\]

where \( T^* \) denotes the adjoint (transpose) of \( T \). We abbreviate \( L_{(G_X, f, T)} \) by \( L_X \) if no confusion arises from the context.

Intuitively, we require that \( f \) is one-one to ensure that \( f \) “embeds” \( X \) in \( G_X \) such that we may perform the shift operation on \( G_X \). Conditions (b) and (c) on \( T \) say that the signal on \( v \in X^0 \) is preserved at its image \( f(v) \) in \( G_X \), while signals on \( V \setminus f(X) \) are formed from an averaging process.

**Lemma 1.** (a) \( L_X \) is symmetric and positive semi-definite.

(b) Constant signals are in the \( 0 \)-eigenspace of \( L_X \). The \( 0 \)-eigenspace \( E_0 \) of \( L_X \) is 1-dimensional if and only if \( G \) is connected.

It implies that the generalized Laplacian \( L_X \) enjoys a few desired properties. In particular, being symmetric permits an orthonormal basis consisting of the eigenvectors of \( L_X \). Therefore, one can devise a Fourier theory analogous to traditional GSP. Moreover, as \( L_X \) is positive semi-definite, we may perform smoothness based learning. The constant vectors belonging to the \( 0 \)-eigenspace is also desirable as it agrees with the intuition that “constant signals are smoothest”.

Lemma 1 asserts that \( L_X \) is indeed very similar to the Laplacian of a graph. However, there are also cases where \( L_X \) cannot be realized as a graph Laplacian, e.g., \( L_X \) may contain positive off-diagonal entries.

In the following, we give an explicit construction of \( L_X \) together with the choice of \((G_X, f, T)\). As an implicit requirement, we would like the construction to recover the usual Laplacian if \( X \) is a graph.

For the simplest case, assume \( X \cong \Delta_n \) is a weighted \( n \)-simplex, i.e., the 1-skeleton \( X^1 \) is a weighted graph. Let \( d_X(v, \cdot) \) be the distance function on \( X \). We label the vertices of \( X \) by \( v_1, \ldots, v_{n+1} \). The graph \( G_X = (V, E) \) is constructed as follows (illustrated in Fig. 3): \( V = \{v_1, \ldots, v_{n+1}, u\} \) with a single additional vertex \( u \), which is understood as the barycenter of \( X \). There is no edge between \( v_i \) and \( v_j \) for any \( i \) and \( j \in \{1, \ldots, n+1\} \). On the other hand, \((v_i, u) \in E \) for each \( i \in \{1, \ldots, n+1\} \).

The edge weight \( w(v_i, u) \) of \((v_i, u) \) is given by:

\[
w(v_i, u) = \left( \frac{n}{2} \right)^{-1} \sum_{v_i \neq v_j \neq v_k \neq v_l} w(v_j, v_k),
\]

where \((v_j, v_k)_{v_i} = d_X(v_i, v_j) + d_X(v_i, v_k) - d_X(v_j, v_k)) / 2 \) and \( n > 1 \). If \( n = 1 \), we change slightly by \( w(v_1, u) = (v_2, v_2)_{v_1} = d(v_1, v_2) \), recovering the usual Laplacian [16].

This construction leads to a canonical choice for \( f \): each \( v_i \in X^0 \) is sent to itself, i.e., \( f(v_i) = v_i \). Moreover, \( T \) is the identity on the \( v_i \) components, and the average of the signals on \( \{v_1, \ldots, v_{n+1}\} \) is given to the \( u \) component. It is easy to check that \((G_X, T, f)\) verifies the conditions of Definition 1. Thus, we have an associated generalized Laplacian \( L_X \).

For a general finite simplicial complex \( X \), we have a decomposition \( X^{\text{max}} \) as the subset of the maximal simplices in \( X \) and the generalized Laplacian

\[
L_X = \sum_{\sigma \in X^{\text{max}}} L_{\sigma},
\]
where the summation is over all maximal simplices of $X$ with appropriate embedding of the vertex indices of $\sigma$ in $X$.

To give some insights of the construction, we notice that for $X \cong \Delta_n$, it is topologically (homotopy) equivalent to a point. Therefore, if we want to approximate $X$ by a graph $G_X$ that preserves this topological property, $G_X$ must be a tree. In addition, if we do not want to break the symmetry of the vertices, the most natural step to do so is to add one additional node (the barycenter) connected to every vertex in the original graph. The edge weights of $G_X$ are chosen to approximate the metric of $X$.

### 4. 2-COMPLEXES AND SIGNAL PROCESSING

In this section, we focus on 2-complexes. For a weighted 2-simplex $X \cong \Delta_2$ with vertices $\{v_1, v_2, v_3\}$, let $d_X(\cdot, \cdot)$ be the edge weight. With the choice of $(G_X, f, T)$ given in the previous section, the edge weights of $G_X$ are

$$a = (v_2, v_3)_{v_1} = \frac{(d_X(v_1, v_3) + d_X(v_1, v_2) - d_X(v_2, v_3))}{2},$$

$$b = (v_1, v_3)_{v_2} = \frac{(d_X(v_2, v_3) + d_X(v_1, v_2) - d_X(v_1, v_3))}{2},$$

$$c = (v_1, v_2)_{v_3} = \frac{(d_X(v_1, v_3) + d_X(v_2, v_3) - d_X(v_1, v_2))}{2},$$

and the generalized Laplacian $L_X$ is given by:

$$L_X = \frac{1}{9} \begin{bmatrix} b + c + 4a & c - 2a - 2b & b - 2a - 2c \\ c - 2a - 2b & a + c + 4b & a - 2b - 2c \\ b - 2a - 2c & a - 2b - 2c & a + b + 4c \end{bmatrix}.$$  

From this explicit formula, we see that for certain $X$, $L_X$ cannot be the usual Laplacian of any graph. For example, if $c > 2a + 2b$ there are some off-diagonal non-negative entries.

If $X$ is a general 2-dimensional simplicial complex, the Laplacian $L_X$ takes contribution from the Laplacian of 2-simplices computed as above and the usual edge Laplacians. We may compare $L_X$ and $L_{X^1}$ as the latter is well-studied. We write $A \preceq B$ if $B - A$ is positive semi-definite.

**Proposition 1.** Suppose $X$ is a finite 2-dimensional simplicial complex with each edge of length 1. Let $k_{\max}$ and $k_{\min}$ be the largest and smallest numbers of 2-simplices that can share a single edge. Then

$$\max \left\{ \frac{k_{\min}}{3}, \frac{1}{3} \right\} \leq L_X \preceq \frac{k_{\max}}{3} L_{X^1}.$$  

For example, if $X$ is a 2D-mesh (triangulation) of a compact 2-manifold, then $k_{\min} = 1$ and $k_{\max} = 2$.

Recall that a filter $F$ is shift invariant w.r.t. $L_{X^1}$ if $F \circ L_{X^1} = L_{X^1} \circ F$. If the graph Laplacian $L_{X^1}$ does not have repeated eigenvalues, then $F$ is shift invariant if and only if $F = P(L_{X^1})$ for some polynomial $P$ of degree less than the number of vertices in $X^1$. The shift invariant family is of particular interest and they are readily estimated. Due to this fact, $L_X$ will be less interesting if it is shift invariant w.r.t. $L_{X^1}$. In general, this does not happen (exemplified in Fig. 4).

![Fig. 3](image)

**Fig. 3.** Graphical illustration of the shape of $G_X$ for $X \cong \Delta_2$ and $X \cong \Delta_3$.

![Fig. 4](image)

**Fig. 4.** In (a), if all the edge weights are the same then $L_X = 1/3 L_{X^1}$ is shift invariant w.r.t. $L_{X^1}$. However, in (b), as long as the 4 distinct red vertices are contained in a graph $G$ (at the center), then $L_X$ is not shift invariant w.r.t. $L_{X^1}$, even if we allow arbitrary positive edge weights.

In a general signal processing problem, the simplicial complex structure might not be given. We now describe a method that approximates the 2-complex structure by using signals on the nodes. If $X^1$ is an unweighted graph, we assign weight 1 to each edge.

We first identify the set $C_X^0$ of all possible 2-simplices. Suppose $X^1$ is given, then triple nodes $(v_1, v_2, v_3)$ belongs to $C_X^0$ if and only if they pair-wise form edges of $X^1$.

Given two non-negative numbers $r_1 \leq r_2$, we define $C_X^0(r_1, r_2)$ to be the subset of $C_X^0$ consisting of triples $(v_1, v_2, v_3)$ whose pairwise edge weights are within the interval $[r_1, r_2]$. Hence, we have the fundamental filtration $\emptyset = C_X^0(0, 0) \subset C_X^0(0, r) \subset C_X^0(0, r') \subset C_X^0(0, \infty) = C_X^0$ for $r \leq r'$. Next, we perform the following steps:

(a) Order all the 2-simplices of $C_X^0$ in a queue $Q$:

(i) Choose $r_0 = 0 \leq r_1 \leq \ldots \leq r_m$ such that $C_X^0 = C_X^0(0, r_m)$. A simplex in $C_X^0(0, r_1)$ is ordered before that in $C_X^0(0, r_{i+1}) \setminus C_X^0(0, r_i)$ (small triangles first).

(ii) We order the 2-simplices of $C_X^0(r_1, r_{i+1})$ in such a way that 2-simplices sharing more edges are ordered later in the queue (see Fig. 5).

(b) Partition $Q$ as a disjoint union $Q = \bigcup_{1 \leq i \leq p} Q_i$ such that their sizes are approximately uniform.
(c) Let \( X_0 \) be \( X^0 \cup X^1 \). For each \( 1 \leq i \leq p \), we construct a 2-complex \( X_i \) by adding the 2-simplices of \( Q_i \) (and the associated edges) to \( X_{i-1} \). We form the associated generalized Laplacians \( L_i = L_{X_i} \).

(d) Approximate the actual Laplacian by using one of \( L_i \).

This step is problem dependent, which in particular relies on the given signal and usually involves an optimization step. We shall be more explicit in Section 5.

Fig. 5. In this example, the blue 2-simplices in (b) are (randomly) ordered first in \( Q \). After which, we have the pink 2-simplices in (c). The green simplices in (d) are ordered last.

Once such a Laplacian is obtained, one may perform signal processing tasks, such as defining Fourier transform, sampling and filtering, similar to traditional GSP.

5. SIMULATION RESULTS

We consider three signal processing tasks: topology inference, signal compression and anomaly detection.

We start with a real social graph \( G \) of size \( n = 500 \) and 6815 pair-wisely connected triples (from [19]). We construct a 2-complex \( X \) by randomly adding 2-simplices for pair-wisely connected triples in \( G \). As a result, \( G = X^1 \) is observed and \( X \) is unobserved. Let \( B = \{ f_1, \ldots, f_n \} \) be an eigenbasis of \( L_X \) arranged according to increasing order of the associated eigenvalues. We randomly generate a set \( S_0 \) of signals from the span of the first \( r_1 \% \) of base signals.

To learn \( X \) from \( S_1 \), we construct \( X_i \) and \( L_{X_i} \) as in Section 4 Step (c) for \( 0 \leq i \leq p = 20 \) where \( X_0 = X^1 = G \) (the spectrum of \( L_{X_i} \) is plotted in Fig. 6). Let \( V_{r_i,i} \) be the matrix whose columns are the first \( r_i \% \) of the eigenvalues of \( L_{X_i} \). Then the estimated simplicial complex \( X_b \) and its Laplacian \( L_{X_b} \) is obtained by solving the optimization problem:

\[
b = \arg \min_{0 \leq i \leq p=20} \sum_{j \in S_1} \| V_{r_1,i} V_{r_1,i}^T - b \|^2.
\]

To test the performance, we generate a set \( S_2 \) from the first \( r_2 \% \) of the base signals in \( B \), considered as a set of compressible signals. We want to estimate the signal compression error of the estimated Laplacian \( L_{X_b} \) as:

\[
err = \sum_{f \in S_2} \| V_{r_2,b} V_{r_2,b}^T - f \|^2.
\]

For comparison, we perform the same estimation on \( L_{X_0} \), for which we do not consider high dimensional structures. On average for different choices of \( X \), as compared to using \( L_{X_0} \), the compression error with \( L_{X_b} \) is reduced by 33.2\% and 40.6\% for \( r_1 \% = r_2 \% = 30 \% \) and 50\%, respectively.

Finally, we introduce anomalies to signals in a new set \( S_3 \) (with \( r_1 \% = 2r_2 \% = 50 \% \)) by perturbing the signal value at one random node. We perform spectrum analysis of the anomalous signals using both \( L_{X_b} \) and \( L_{X_b} \), again as a comparison between with or without high dimensional structures.

Two typical examples of the spectral plots are shown in Figure 7. We see that using \( L_{X_b} \) (red), the anomalous behavior is more easily detected by inspecting high frequency portions.

Fig. 6. The plot shows the eigenvalue distribution of \( L_{X_b} = L_G \) to \( L_{X_20} \), from bottom to top. The spectrum tends to shift to the left, which might show a “more connected” structure.

Fig. 7. We have two sets of high frequency component plots. For each set, the left figures (blue) are normal and abnormal plots of \( L_{X_0} \), and the right figures (red) are the plots of \( L_{X_b} \). The anomalous behavior is more visible for \( L_{X_b} \).

6. CONCLUSIONS

In this paper, we have proposed a signal processing framework for signals on simplicial complexes. To do so, we introduced a general way to construct a Laplacian matrix on a space, which may not be a graph. After which, signal processing follows in the way similar to traditional GSP. We test the framework with synthetic data. For future work, we shall apply the approach to more real world problems.
7. REFERENCES

[1] David I Shuman, Sunil K Narang, Pascal Frossard, Antonio Ortega, and Pierre 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] Aliaksei Sandryhaila and José M. F. Moura, “Discrete signal processing on graphs,” *IEEE Trans. Signal Process.*, vol. 61, no. 7, pp. 1644–1656, April 2013.

[3] Aliaksei Sandryhaila and José M. F. Moura, “Big data analysis with signal processing on graphs: Representation and processing of massive data sets with irregular structure,” *IEEE Signal Process. Mag.*, vol. 31, no. 5, pp. 80–90, Sept 2014.

[4] Xiaowen Dong, Dorina Thanou, Pascal Frossard, and Pierre Vandergheynst, “Learning laplacian matrix in smooth graph signal representations,” *IEEE Transactions on Signal Processing*, vol. 64, no. 23, pp. 6160–6173, 2016.

[5] Hilmi E. Egilmez, Eduardo Pavez, and Antonio Ortega, “Graph learning from data under laplacian and structural constraints,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 6, pp. 825–841, 2017.

[6] Francesco Grassi, Andreas Loukas, Nathanaël Perraudin, and Benjamin Ricaud, “A time-vertex signal processing framework: Scalable processing and meaningful representations for time-series on graphs,” *IEEE Trans. Signal Process.*, vol. 66, no. 3, pp. 817–829, Feb 2018.

[7] Antonio Ortega, Pascal Frossard, Jelena Kovačević, José MF Moura, and Pierre Vandergheynst, “Graph signal processing: Overview, challenges, and applications,” *Proceedings of the IEEE*, vol. 106, no. 5, pp. 808–828, 2018.

[8] Feng Ji and Wee Peng Tay, “A Hilbert space theory of generalized graph signal processing,” *IEEE Trans. Signal Process.*, vol. 67, no. 24, pp. 6188 – 6203, Dec. 2019.

[9] M. Defferrard, X. Bresson, and P. Vandergheynst, “Convolutional neural networks on graphs with fast localized spectral filtering,” in *Advances in Neural Inform. Process. Syst.*, USA, 2016, pp. 3844–3852.

[10] Thomas N. Kipf and Max Welling, “Semi-supervised classification with graph convolutional networks,” *arXiv preprint arXiv:1609.02907*, 2016.

[11] Ruoyu Li, Sheng Wang, Feiyun Zhu, and Junzhou Huang, “Adaptive graph convolutional neural networks,” in *Thirty-second AAAI conference on artificial intelligence*, 2018.

[12] Sergio Barbarossa and Stefania Sardellitti, “Topological signal processing over simplicial complexes,” *arXiv preprint arXiv:1907.11577*, 2019.

[13] Sergio Barbarossa and Mikhail Tsitsvero, “An introduction to hypergraph signal processing,” in *IEEE Int. Conf. Acoustics, Speech and Signal Process.*, 2016, pp. 6425–6429.

[14] Markus Puschel, “A discrete signal processing framework for meet/join lattices with applications to hypergraphs and trees,” in *IEEE Int. Conf. Acoustics, Speech and Signal Process.*, 2019, pp. 5371–5375.

[15] Songyang Zhang, Zhi Ding, and Shuguang Cui, “Introducing hypergraph signal processing: theoretical foundation and practical applications,” *arXiv preprint arXiv:1907.09203*, 2019.

[16] Feng Ji, Giacomo Kahn, and Wee Peng Tay, “Signal processing on simplicial complexes,” Work in progress, available at giacomo.kahn.science/resources/SPonSC.pdf, 2020.

[17] Edwin H Spanier, *Algebraic topology*, Springer Science & Business Media, 1989.

[18] Allen Hatcher, *Algebraic Topology*, Cambridge University Press, 2002.

[19] Bryan Klimt and Yiming Yang, “Introducing the enron corpus,” in *CEAS*, 2004.