Density of States of Hodge Laplacians: Decomposition effects and Sparsification of SC

Tony Savostianov [1]

[1]*RWTH, Aachen , email:* a.s.savostyanov@gmail.com

Abstract:

Keywords: density of states, simplicial complexes, sparsification, generalized effective resistance

I Introduction

Well, it would be nice to have one.
You know, somewhere here, maybe...

II Graphs and Simplicial Complexes

Simplicial complex $\mathcal{K}$ is a higher-order generalization of the classical graph model with pair-wise interactions with a rich set of topological descriptors. Specifically, let $\mathcal{V}_0(\mathcal{K}) = \{v_1, \dots v_{m_0}\}$ be a set of nodes; then $\mathcal{K}$ is a collection of subsets (simplices) σ of nodes from $\mathcal{V}_0(\mathcal{K})$ such that all the subsets of σ are simplices in $\mathcal{K}$ too. We refer to a simplex made out of $k+1$ nodes $\sigma = [v_{i_1}, \dots v_{i_{k+1}}]$ as being of order $k$the set of all the simplices of order k in the complex $\mathcal{K}$ is denoted by $\mathcal{V}_k(\mathcal{K})$. Thus, $\mathcal{V}_0(\mathcal{K})$ are the vertices of $\mathcal{K}$, $\mathcal{V}_1(\mathcal{K})$ are edges between pairs of vertices, $\mathcal{V}_2(\mathcal{K})$ triangles connecting three vertices, and so on. We let $m_k = |\mathcal{V}_k(\mathcal{K})|$ denote the cardinality of $\mathcal{V}_k(\mathcal{K})$.

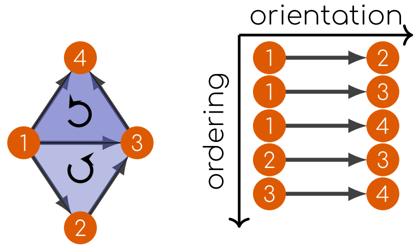
Each set of simplices $\mathcal{V}_k(\mathcal{K}) = \left\{\sigma_1, \dots \sigma_{m_k}\right\}$ induces a linear space of formal sums over the simplicies $C_k(\mathcal{K}) = \left\{\sum_{i=1}^{m_k} \alpha_i \sigma_i \mid \alpha_i \in \mathbb{R}\right\}$ referred to as *chain space*; in particular, $C_0(\mathcal{K})$ is known as the space of vertex states and $C_1(\mathcal{K})$ as the space of edge flows. Simplices of different orders are related through the boundaries operators ∂_k mapping the simplex to its boundary; formally, $\partial_k : C_k(\mathcal{K}) \mapsto C_{k-1}(\mathcal{K})$ is defined through the alternating sum:

$$\partial_k [v_1, v_2, \dots v_k] = \sum_{i=1}^{k} (-1)^{i-1} [v_1, \dots v_{i-1}, v_{i+1}, \dots v_k]$$

By fixing an ordering for $\mathcal{V}_k(\mathcal{K})$ we can fix a canonical basis for $C_k(\mathcal{K})$ and represent each boundary operator as a matrix $B_k \in \mathrm{Mat}_{m_{k-1} \times m_k}$ with exactly k nonzero entries per each column, being either $+1$ or -1. For these matrices, the fundamental property of topology holds: *the boundary of the boundary is zero*, [Lim20, Thm. 5.7]:

$$B_k B_{k+1} = 0 \qquad \text{(Eqn. 1)}$$

The matrix representation B_k of the boundary operator ∂_k requires fixing an ordering of the simplices in $\mathcal{V}_k(\mathcal{K})$ and $\mathcal{V}_{k-1}(\mathcal{K})$. As it will be particularly relevant for the purpose of this work, we emphasize that we order triangles and edges as follows: triangles in $\mathcal{V}_2(\mathcal{K})$ are oriented in such a way that the first edge (in terms of the ordering of $\mathcal{V}_1(\mathcal{K})$) in each triangle is positively acted upon by B_2, i.e. the first non-zero entry in each column of B_2 is $+1$, Figure .1.



$$
\begin{aligned}
B_2[1,2,3] &= \overbrace{(+1)[1,2]}^{\text{1st in order}} + (-1)[1,3] + (+1)[2,3] \\
B_2[1,3,4] &= \underbrace{(+1)[1,3]}_{\text{1st in order}} + (-1)[1,4] + (+1)[3,4]
\end{aligned}
$$

Figure .1: Example of the simplicial complex with ordering and orientation: nodes from $\mathcal{V}_0(\mathcal{K})$ in orange, triangles from $\mathcal{V}_2(\mathcal{K})$ in blue. Orientation of edges and triangles is shown by arrows; the action of B_2 operator is given for both triangles.

The following definitions introduce the fundamental concepts of k-th homology group and k-th order Laplacian. See [Lim20] e.g. for more details.

Def. 1 **(Homology group and higher-order Laplacian)** Since $\operatorname{im} B_{k+1} \subset \ker B_k$, the quotient space $\mathcal{H}_k = {}^{\ker B_k}/_{\operatorname{im} B_{k+1}}$, known as k-th homology group, is correctly defined and the following isomorphisms hold

$$\mathcal{H}_k \cong \ker B_k \cap \ker B_{k+1}^\top \cong \ker\left(B_k^\top B_k + B_{k+1}B_{k+1}^\top\right).$$

The matrix $L_k = B_k^\top B_k + B_{k+1}B_{k+1}^\top$ is called the k-th order **Laplacian operator**; the two terms $L_k^\downarrow = B_k^\top B_k$ and $L_k^\uparrow = B_{k+1}B_{k+1}^\top$ are referred to as the **down-Laplacian** and the **up-Laplacian**, respectively.

The homology group $\mathcal{H}_k$ describes the k-th topology of the simplicial complex $\mathcal{K}$: $\beta_k = \dim \mathcal{H}_k = \dim\ker L_k$ coincides exactly with the number of k-dimensional holes in the complex, known as the **k-th Betti number**. In the case $k = 0$, the operator $L_0 = L_0^\uparrow$ is exactly the classical graph Laplacian whose kernel corresponds to the *connected components* of the graph, while $L_0^\downarrow = 0$. For $k = 1$ and $k = 2$, the elements of $\ker L_1$ and $\ker L_2$ describe the simplex 1-dimensional holes and voids respectively, and are frequently used in the analysis of trajectory data, [SBH+20, BGL16].

Although more frequently found in their purely combinatorial form, the definitions of simplicial complexes, homology groups, and higher-order Laplacians admit a generalization to the weighted case. For the sake of generality, in the rest of the work, we use the following notion of weighted boundary operators (and thus weighted simplicial complexes), as considered in e.g. [GST23].

Def. 2 **(Weighted and normalised boundary matrices)** For *weight functions* $w_k : \mathcal{V}_k(\mathcal{K}) \mapsto \mathbb{R}_+ \cup \{0\}$, define the diagonal weight matrix $W_k \in \mathrm{Mat}_{m_k \times m_k}$ as $(W_k)_{ii} = \sqrt{w_k(\sigma_i)}$. Then the weighting scheme for the boundary operators upholding the Hodge algebras 1 is given by:

$$B_k \mapsto W_{k-1}^{-1} B_k W_k \tag{Eqn. 2}$$

Note that, with the weighting scheme Equation (2), the dimensionality of the homology group is preserved, $\dim\ker L_k = \dim\ker \widehat{L}_k$, [GST23] as well as the fundamental property of topology Equation (1).

III Density of States and Hodge Decomposition

add some motivation for this discussion/transition

Full spectral decomposition of $L_k^{\uparrow}$ requires $\mathcal{O}(m_k^3)$ (or, more precisely, $\mathcal{O}(m_k^{\omega})$ where ω is the exponent of matrix multiplication, [BGVKS23]); accounting for the sparsity of $L_k^{\uparrow}$, the actual computation time is even lower. As a result, even so we are going to use $\mathcal{O}(m_k^3)$ as a reference point to beat, the actual execution time should be carefully compared.

Instead of the direct computation, one can exploit spectral density functions known as **density of states** which encompass all the spectral information. Specifically,

Def. 3 **(Density of States)** Fro a given symmetric matrix $A = Q\Lambda Q^\top$ with $Q^\top Q = I$ and diagonal $\Lambda = \mathrm{diag}\,(\lambda_1, \dots \lambda_n)$, the **spectral density** or **density of states** (DoS)

$$\mu(\lambda \mid A) = \frac{1}{n} \sum_{i=1}^{n} \delta\left(\lambda - \lambda_i\right) \tag{Eqn. 3}$$

Additionally, let $\mathbf{q}_i$ be a corresponding unit eigenvector of A (such that $A\mathbf{q}_i = \lambda_i \mathbf{q}_i$ and $Q = (\mathbf{q}_1 \mid \mathbf{q}_2 \mid \cdots \mid \mathbf{q}_n)$); then one can define a set of local (entry-wise) densities of states (**LDoS**):

$$\mu_k(\lambda \mid A) = \sum_{i=1}^{n} \left|\mathbf{e}_k^\top \mathbf{q}_i\right|^2 \delta\left(\lambda - \lambda_i\right) \tag{Eqn. 4}$$

with $\mathbf{e}_k$ being the corresponding versor.

Note that DoS $\mu(\lambda \mid A)$ only describes the spectrum $\sigma(A)$ while the family of LDoS $\{\mu_k(\lambda \mid A)\}$ contain additionally eigenvectors $\mathbf{q}_i$. The exact computation of either DoS or LDoS still requires the full spectral decomposition; instead, one can try to approximate densities of states via Kernel Polynomial Method (KPM).

I. Chebyshev Approximation of DoS/LDoS

Let us assume we transition from symmetric A to a symmetric H such that $\sigma(H) \subseteq [-1, 1]$, e.g. by $H = \frac{2}{\lambda_{\max}(A)} A - I$. This allows us to assume that $\operatorname{supp} \mu(\lambda \mid H) \subseteq [-1, 1]$ and it can be decomposed into an orthonormal basis on $[-1, 1]$.

Specifically, let $T_m(x)$ be a family of Chebyshev polynomials of the first kind iff:

$$T_0(x) = 1, \quad T_1(x) = x, \qquad T_{m+1}(x) = 2xT_m(x) - T_{m-1}(x) \tag{Eqn. 5}$$

Polymonials $T_m(x)$ are famously orthogonal on $[-1, 1]$ in L_2-scalar product with the weight function $w(x) = \frac{2}{(1+\delta_{0m})\pi\sqrt{1-x^2}}$. Then, $T^*_m(x) = w_m(x)T_m(x)$ form a dual Chebyshev basis, and one can write an expansion:

figure out how to write here the problem with $m = 0$ for the orthonormality

$$\begin{aligned} \mu(\lambda \mid H) &= \sum_{m=0}^{\infty} d_m T^*_m(\lambda) \\ \mu_k(\lambda \mid H) &= \sum_{m=0}^{\infty} d_{mk} T^*_m(\lambda) \end{aligned} \tag{Eqn. 6}$$

Note that since $T_m(x)$ are orthogonal to $T^*_m(x)$, so

$$\begin{aligned} d_m &= \int T_m(\lambda)\mu(\lambda \mid H) d\lambda = \frac{1}{n}\sum_{i=1}^{n} T_m(\lambda_i) = \frac{1}{n}\operatorname{tr}(T_m(H)) \\ d_{mk} &= \int T_m(\lambda)\mu_k(\lambda \mid H) d\lambda = \sum_{i=1}^{n} \left|\mathbf{e}_k^\top \mathbf{q}_i\right|^2 T_m(\lambda_i) = [T_m(H)]_{kk} \end{aligned} \tag{Eqn. 7}$$

with

$$\begin{aligned} \mathbb{E}\left[\mathbf{z}^\top H\mathbf{z}\right] &= \operatorname{tr}(H) \approx \frac{1}{N_z}\sum_{j=1}^{N_z} Z_j^\top H Z_j \\ \mathbb{E}\left[\mathbf{z} \odot H\mathbf{z}\right] &= \operatorname{diag}(H) \approx \frac{1}{N_z}\sum_{j=1}^{N_z} Z_j \odot H Z_j \end{aligned} \tag{Eqn. 8}$$

where each entry in $\mathbf{z}$ is i.i.d with zero mean and unit variance.

Note that due to the recurrent definition, $d_m = \frac{1}{n}\mathbb{E}\left[\mathbf{z}^\top T_m(H)\mathbf{z}\right] = \frac{1}{n}\mathbb{E}\left[\mathbf{z}^\top \left(2HT_{m-1}(H) - T_{m-2}(H)\right)\mathbf{z}\right]$, the computation of d_m requires only one new `matvec` computation for $HT_{m-1}(H)\mathbf{z}$ since estimations for $T_{m-1}(H)\mathbf{z}$ and $T_{m-2}(H)\mathbf{z}$ are inherited from d_{m-1} and d_{m-2}. As a result, the computational cost of each new moment d_m is $\mathcal{O}(\mathtt{nnz}(H))$, and, assuming we use M moments in the decomposition, the overall complexity is $\mathcal{O}\left(N_z M \mathtt{nnz}(H)\right)$.

In the case of L_k, we get $\mathcal{O}\left(N_z M(km_k + (k+1)m_{k+1})\right)$ either for DoS of LDoS.

II. Spike problems and filtration

Approximation via KPM has an unfortunate down-side: since it is a basis decomposition, a local failure (e.g. in the neibourhood of a a specific λ) affects the approximation quality everywhere. Moreover, since $T^*_m(x)$ are mostly polynomial, spikes in the spectral densities (i.e. associated with eigenvalues of high multiplicity) are poorly approximated in general.

III. Effects of the Hodge Decomposition

IV Sparsification of Simplicial Complexes

Simplicial complex $\mathcal{K}$ typically has quite an intrinsic structure of the associated Laplacian operators L_k which becomes more and more complicated as $\mathcal{K}$ becomes denser in the sense of complexes of order $\mathcal{K}$. Instead, one can ask a simplifying question: can one find a sparser simplicial complex $\mathcal{L}$ with a spectrally close operator $L_k(\mathcal{L})$?

The overall idea of the sparsification is built upon the idea that a sparsifier can be subsampled from a given simplificial copmlex with high probability. Specifically:

Def. 4 **(Spectral Approximation)** The Hermitian matrix A is called **spectrally ε-close** to the Hermitian matrix B, $A \underset{\varepsilon}{\approx} B$, if $(1-\varepsilon)B \preceq A \preceq (1+\varepsilon)B$, where $\succeq$ is the partial ordering induced by the positive definite cone, i.e. $A \succeq B$ if $A - B$ is positive semi-definite.

Rem III.1 Note that if $A \underset{\varepsilon}{\approx} B$, then one can directly bound the distance $\|\mathbf{x}_A - \mathbf{x}_B\|$, where $A\mathbf{x}_A = \mathbf{f}$ and $B\mathbf{x}_B = \mathbf{f}$. Indeed, $\mathbf{x}_B = B^{-1}A\mathbf{x}_A$ and $\|\mathbf{x}_A - \mathbf{x}_B\| = \left\| \left(I - B^{-1}A\right)\mathbf{x}_A \right\| \le \|I - B^{-1}A\|\|\mathbf{x}_A\| \le \|B^{-1}\| \cdot \|B - A\| \cdot \|\mathbf{x}_A\| \le \varepsilon\|B^{-1}\| \cdot \|B\| \cdot \|\mathbf{x}_A\| = \varepsilon\kappa(B)\|\mathbf{x}_A\|$, with $\kappa(B)$ being the condition number of B. Thus, the relative error is controlled by the quality of the approximation ε and the condition number $\kappa(B)$.

Th IV.III.1 **(Simplicial Sparsification, [OPW22])** Let $\mathcal{K}$ be a simplicial complex restricted to its p-skeleton, $\mathcal{K} = \bigcup_{i=0}^{p} \mathcal{V}_i(\mathcal{K})$. Let $L_k^{\uparrow}(\mathcal{K})$ be its k-th up-Laplacian and let $m_k = |\mathcal{V}_k(\mathcal{K})|$. For any $\varepsilon > 0$, a sparse simplicial complex $\mathcal{L}$ can be sampled as follows:

(1) compute the probability measure $\mathbf{p}$ on $\mathcal{V}_{k+1}(\mathcal{K})$ proportional to the generalized resistance vector $\mathbf{r} = \operatorname{diag}\left(B_{k+1}^\top (L_k^{\uparrow})^\dagger B_{k+1}\right)$, where $\operatorname{diag}(A)$ denotes the vector of the diagonal entries of A;

(2) sample q simplices τ_i from $\mathcal{V}_{k+1}(\mathcal{K})$ according to the probability measure $\mathbf{p}$, where q is chosen so that $q(m_k) \ge 9C^2 m_k \log(m_k/\varepsilon)$, for some absolute constant $C > 0$;

(3) form a sparse simplicial complex $\mathcal{L}$ with all the sampled simplexes of order k and all its faces with the weight $\frac{w_{k+1}(\tau_i)}{q(m_k)\mathbf{p}(\tau_i)}$; weights of repeated simplices are accumulated.

Then, with probability at least $1/2$, the up-Laplacian of the sparsifier $\mathcal{L}$ is ε-close to the original one, i.e. it holds $L_k^{\uparrow}(\mathcal{L}) \underset{\varepsilon}{\approx} L_k^{\uparrow}(\mathcal{K})$.

The bottleneck of the subsampling above is the construction of the appropriate measure $\mathbf{p}$ or, more precisely, generalized effective resistance $\mathbf{r}$. Indeed, one needs a fast pseudo-inverse operator $\left(L_k^{\uparrow}\right)^{\dagger}$ in order to compute $\mathbf{r}$.

Rem III.2 **(Sensitivity of the Sparsification vis-a-vis sampling measure p)** Compare ε with $\frac{1}{m_2}$: if below, you are fiiiiiiine
INSERT FIGURE HERE

we need to assume somewhere $W_k = I$ for simplicity and do not forget about it

Th IV.III.2 **(GER through DoS)** For a given simplicial complex mcK with the k-th order up-Laplacian $L_k^{\uparrow} = B_{k+1} W_{k+1}^2 B_{k+1}^\top$, a generalized effective resistance r can be computed through family of local densities of states $\{\mu_i(\lambda \mid L_{k+1}^{\downarrow})\}$:

$$\mathbf{r}_i = \int_{\mathbb{R}} (1 - \mathbb{1}_0(\lambda)) \mu_i(\lambda \mid L_{k+1}^{\downarrow}) d\lambda \quad \text{(Eqn. 9)}$$

Proof Let $B_{k+1} W_{k+1} = U S V^\top$ where S is diagonal and invertible and both U and V are orthogonal (so it is a truncated SVD decomposition of $B_{k+1} W_{k+1}$ matrix with eliminated obsolete kernel). Then:

$$(L_k^{\uparrow})^{\dagger} = \left(B_{k+1} W_{k+1}^2 B_{k+1}^\top \right)^{\dagger} = \left(U S^2 U^\top \right)^{\dagger} = U S^{-2} U^\top \quad \text{(Eqn. 10)}$$

$$\begin{aligned} \mathbf{r} &= \operatorname{diag} \left(W_{k+1} B_{k+1}^\top (L_k^{\uparrow})^{\dagger} B_{k+1} W_{k+1} \right) = \\ &= \operatorname{diag} \left(V S U^\top U S^{-2} U^\top U S V^\top \right) = \operatorname{diag} \left(V V^\top \right) \end{aligned} \quad \text{(Eqn. 11)}$$

As a result, $\mathbf{r}_i = \|V_{i\cdot}\|^2 = \sum_j |v_{ij}|^2$, so the i-th entry of the resistance is defined by the sum of square of i-th components of eigenvectors $\mathbf{v}_j$ of $L_{k+1}^{\downarrow} = W_{k+1} B_{k+1}^\top B_{k+1} W_{k+1}$ operator where $\mathbf{v}_j \perp \ker L_{k+1}^{\downarrow}$.
Note that

$$\mu_i(\lambda \mid L_{k+1}^{\downarrow}) = \sum_{j=1}^{m_{k+1}} \left| \mathbf{e}_i^\top \mathbf{q}_j \right|^2 \delta\left(\lambda - \lambda_j\right) = \sum_{j=1}^{m_{k+1}} |q_{ij}|^2 \delta\left(\lambda - \lambda_j\right) \quad \text{(Eqn. 12)}$$

so

$$\begin{aligned} \mathbf{r}_i &= \|V_{i\cdot}\|^2 = \sum_j |v_{ij}|^2 = \int_{\mathbb{R}\setminus\{0\}} \sum_{j=1}^{m_{k+1}} |q_{ij}|^2 \delta\left(\lambda - \lambda_j\right) d\lambda = \\ &= \int_{\mathbb{R}\setminus\{0\}} \mu_i(\lambda \mid L_{k+1}^{\downarrow}) d\lambda = \int_{\mathbb{R}} (1 - \mathbb{1}_0(\lambda)) \mu_i(\lambda \mid L_{k+1}^{\downarrow}) d\lambda \end{aligned} \quad \text{(Eqn. 13)}$$

■

- check for the exact case
- error propagation

IV References

[BGL16] Austin R Benson, David F Gleich, and Jure Leskovec. Higher-order organization of complex networks. *Science*, 353(6295):163–166, 2016.

[BGVKS23] Jess Banks, Jorge Garza-Vargas, Archit Kulkarni, and Nikhil Srivastava. Pseudospectral shattering, the sign function, and diagonalization in nearly matrix multiplication time. *Foundations of computational mathematics*, 23(6):1959–2047, 2023.

[DBB19] Kun Dong, Austin R Benson, and David Bindel. Network density of states. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1152–1161, 2019.

[GST23] Nicola Guglielmi, Anton Savostianov, and Francesco Tudisco. Quantifying the Structural Stability of Simplicial Homology. *Journal of Scientific Computing*, 97(1):2, August 2023.

[Lim20] Lek-Heng Lim. Hodge laplacians on graphs. *Siam Review*, 62(3):685–715, 2020.

[OPW22] Braxton Osting, Sourabh Palande, and Bei Wang. Spectral sparsification of simplicial complexes for clustering and label propagation. *Journal of computational geometry*, 11(1), 2022.

[SBH+20] Michael T Schaub, Austin R Benson, Paul Horn, Gabor Lippner, and Ali Jadbabaie. Random walks on simplicial complexes and the normalized Hodge 1-Laplacian. *SIAM Review*, 62(2):353–391, 2020.