Geometric techniques in topological data analysis

Toward persistent Hodge Theory

Thomas Sachen

Areté, 2023

Outline

Table of Contents

Differential forms on real manifolds.

Definition

A real n-manifold is a topological space X that is locally homeomorphic to $\mathbb{R}^n$.

Differential forms on real manifolds.

Definition

A real n-manifold is a topological space X that is locally homeomorphic to $\mathbb{R}^n$.

Remark

Furthermore, X is smooth if, for any $x_1, x_2 \in X$ with associated open neighborhoods U_i and homeomorphisms $\phi_i : U_i \to \mathbb{R}^n$, the composition

$$\phi_2 \circ \phi_1^{-1} : \mathbb{R}^n \to \mathbb{R}^n$$

is infinitely differentiable on $\phi_1(U_1 \cap U_2) = \phi_2(U_1 \cap U_2)$.

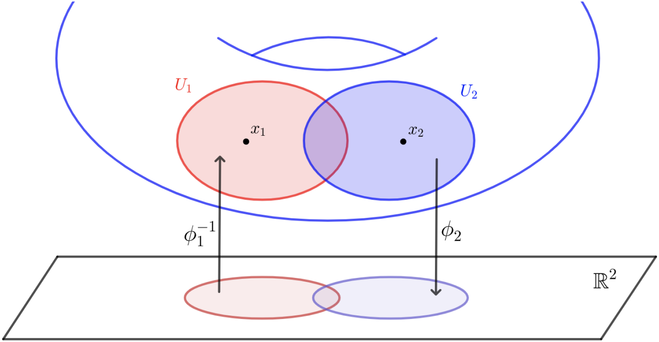


Figure 1: We require that the composition $\phi_2 \circ \phi_1^{-1}$ is smooth.

Definition

A differential n-form α on a smooth manifold X is a section $\alpha : X \to \bigwedge^n T^*X = \bigwedge^n \mathrm{Hom}(T^*X, \mathbb{R})$ of the cotangent bundle over X.

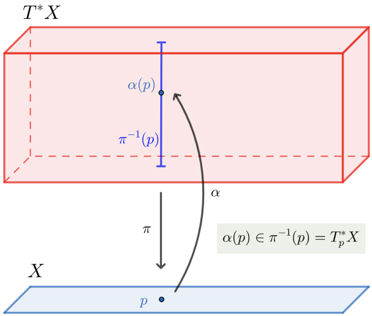


Figure 2: For any $p \in X$, it is the assignment $\alpha : p \mapsto (\phi_p : T_pX \to \mathbb{R})$.

Hodge Theory

Let X be a compact smooth manifold, and let $\Omega^k(X)$ be the vector space of differential k-forms on X, with $\alpha \sim \beta$ if $\alpha - \beta$ is exact.

Definition

Denote by $\mathrm{H}^k_{\mathrm{dR}}(X)$ the k^{th} homology of the chain complex induced by the exterior derivative $d_i : \Omega^i(X) \to \Omega^{i+1}(X)$:

$$\mathcal{O}(X) \xrightarrow{d_0} \Omega^1(X) \xrightarrow{d_1} \dots \xrightarrow{d_{i-1}} \Omega^i(X) \xrightarrow{d_i} \dots$$

$\mathrm{H}^k_{\mathrm{dR}}(X)$ is the k^{th} de Rham cohomology of X.

Hodge Theory

Let X be a compact smooth manifold, and let $\Omega^k(X)$ be the vector space of differential k-forms on X, with $\alpha \sim \beta$ if $\alpha - \beta$ is exact.

Definition

Denote by $\mathrm{H}^k_{\mathrm{dR}}(X)$ the k^{th} homology of the chain complex induced by the exterior derivative $d_i : \Omega^i(X) \to \Omega^{i+1}(X)$:

$$\mathcal{O}(X) \xrightarrow{d_0} \Omega^1(X) \xrightarrow{d_1} \dots \xrightarrow{d_{i-1}} \Omega^i(X) \xrightarrow{d_i} \dots$$

$\mathrm{H}^k_{\mathrm{dR}}(X)$ is the k^{th} de Rham cohomology of X.

That is, $\mathrm{H}^k_{\mathrm{dR}}(X) = \frac{\ker(d:\Omega^k(X)\to\Omega^{k+1}(X))}{\mathrm{im}(d:\Omega^{k-1}(X)\to\Omega^k(X))}$.

Hodge theory (cont.)

Remark

There is also a *codifferential* operator,

$$\partial_i : \Omega^i(X) \to \Omega^{i-1}(X),$$

defined as the formal adjoint of d.

Definition

$\triangle_i = (d_i + \partial_i)^2$ is called the i^{rh} Hodge Laplacian. A differential form α annihilated by the k^{rh} Hodge Laplacian is a harmonic k-form.

The heart of Hodge theory

Theorem (de Rham's Theorem)

Let X be a smooth manifold. Then for each $k \in \mathbb{Z}_{\geq 0}$, there is an isomorphism

$$\rho_k : \mathrm{H}^k_{\mathrm{dR}}(X) \xrightarrow{\sim} \mathrm{H}^k(X, \mathbb{R})$$

induced by integration of differential forms:

$$\rho_k : [\alpha] \mapsto \left(C \mapsto \int_C \alpha \right).$$

Theorem (Harmonic representatives)

Each de Rham cohomology class $[\alpha] \in \mathrm{H}^k_{\mathrm{dR}}(X)$ contains a unique harmonic form h_α.

The heart of Hodge theory

Combining the above theorems with the fact that $\dim \mathrm{H}^k(X, \mathbb{R}) = \dim \mathrm{H}_k(X, \mathbb{R})$, we obtain a bijection

$$\left\{ \begin{matrix} \text{harmonic } k\text{−forms} \\ \text{on } X \end{matrix} \right\} \overset{1:1}{\longleftrightarrow} \left\{ \begin{matrix} \text{generators of} \\ \mathrm{H}_k(X, \mathbb{R}) \end{matrix} \right\}$$

The heart of Hodge theory

Combining the above theorems with the fact that $\dim \mathrm{H}^k(X, \mathbb{R}) = \dim \mathrm{H}_k(X, \mathbb{R})$, we obtain a bijection

$$\left\{ \begin{array}{c} \text{harmonic } k\text{−forms} \\ \text{on } X \end{array} \right\} \overset{1:1}{\longleftrightarrow} \left\{ \begin{array}{c} \text{generators of} \\ \mathrm{H}_k(X, \mathbb{R}) \end{array} \right\}$$

Proposition

The multiplicity of the 0-eigenvalue of the k^{th} Hodge Laplacian on X counts generators of the k^{th} simplicial homology of X.

p-adic Hodge Theory.

When X is instead a compact Kähler manifold, we have the canonical Hodge decomposition

$$H^k(X, \mathbb{C}) = \bigoplus_{p+q=k} H^{p,q} \left(\xrightarrow{\sim} \bigoplus_{p+q=k} \mathscr{H}^{p,q}(X) \xrightarrow{\sim} \bigoplus_{p+q=k} H^q(X, \Omega_X^p) \right)$$

p-adic Hodge Theory.

The period isomorphism can be defined for general schemes over $\mathbb{C}$ as follows.

Theorem

Let X/k a k-scheme, with $\iota : k \hookrightarrow \mathbb{C}$ an embedding. Then there is an isomorphism

$$H^*(X, \Omega^\bullet_{X/k}) \otimes_k \mathbb{C} \xrightarrow{\sim} H^*(X_\iota^{an}, \mathbb{Q}) \otimes_{\mathbb{Q}} \mathbb{C}$$

depending on ι. Furthermore if X/k is proper and smooth, this isomorphism is compatible with the Hodge Filtration induced by the naive filtrations on the de Rham complexes.

p-adic Hodge Theory.

There is actually an analogous version of Hodge theory adapted non-archimedean (*e.g.* p-adic) settings, with a proof in the same spirit.

Theorem

Let X/k be a smooth k-variety with smooth normal crossing compactification, with k a p-adic field. Then there is a comparison isomorphism

$$\rho : H^*_{\mathrm{dR}}(X) \otimes_k B_{\mathrm{dR}} \xrightarrow{\sim} H^*_{\text{ét}}(X, \mathbb{Z}_p) \otimes_{\mathbb{Z}_p} B_{\mathrm{dR}}$$

between de Rham and étale (Weil) cohomology theories.

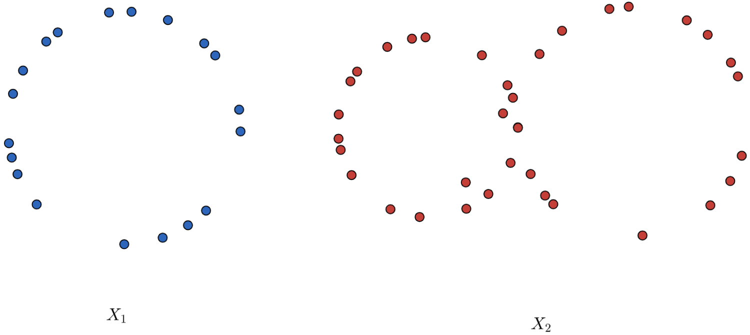


Figure 3: Two point clouds. How can we compare their topological features?

Definition

A finite union of points $X = \{x_i\}_{i=1}^{N}$ in $\mathbb{R}^n$ is a point cloud.

Definition

The union of balls $X_r = \bigcup B_{x_i}(r)$ is the r-augmented point cloud.

Definition

A finite union of points $X = \{x_i\}_{i=1}^N$ in $\mathbb{R}^n$ is a point cloud.

Definition

The union of balls $X_r = \bigcup B_{x_i}(r)$ is the r-augmented point cloud.

Remark

A sequence $0 = r_0 \leq r_1 \leq \dots$ of real numbers induces a filtration of manifolds

$$X_0 \hookrightarrow X_1 \hookrightarrow X_2 \hookrightarrow \dots$$

which, for each $k \in \mathbb{Z}_{\geq 0}$, induces a sequence of vector spaces

$$\mathcal{M}_k := H_k(X_0) \xrightarrow{\phi_1} H_k(X_1) \xrightarrow{\phi_2} H_k(X_2) \xrightarrow{\phi_3} \dots$$

called the persistence module of X (with respect to the r_i).

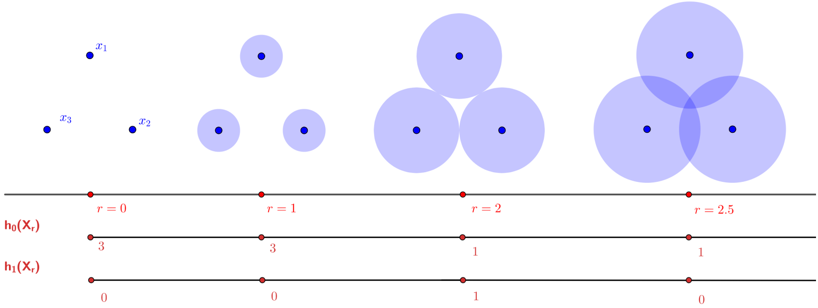


Figure 4: The augmented point clouds X_r and their Betti numbers

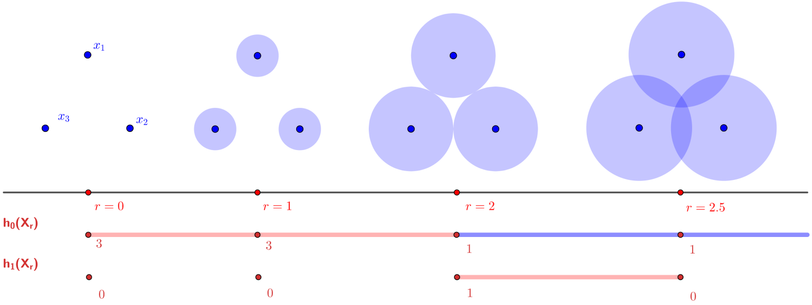


Figure 5: The associated persistence barcodes.

Persistence barcodes

Theorem (Structure theorem)

Each persistence module $\mathcal{M}_k$ admits a unique decomposition as a finite direct sum of interval modules

$$\mathcal{M}_k = \bigoplus_{r_i} \mathbb{I}_{r_1,r_2}$$

where, for $r_1 \leq r_2$ indices of the filtration, $\mathbb{I}_{r_1,r_2}$ is a persistence module of the form

$$\mathbb{I}_{r_1,r_2} = \underbrace{0 \to \cdots \to 0}_{r_1} \to \underbrace{k \xrightarrow{\mathrm{Id}} \ldots \xrightarrow{\mathrm{Id}} k}_{r_2 - r_1} \to 0 \to \ldots,$$

Definition

The k^{th} persistence barcode $\mathrm{Bar}_k(X)$ of $(X, \{r_i\})$ is the multiset of intervals

$$\mathrm{Bar}_k(X) = \{(r_{i_1}, r_{j_1}), (r_{i_2}, r_{j_2}), \dots\},$$

where the (r_i, r_j) are the indices appearing in the decomposition

$$\mathcal{M}_k = \mathbb{I}_{r_{i_1}, r_{j_1}} \oplus \mathbb{I}_{r_{i_2}, r_{j_2}} \oplus \dots$$

. The r_i and r_j birth times and death times, respectively.

Definition

The k^{th} persistence barcode $\mathrm{Bar}_k(X)$ of $(X, \{r_i\})$ is the multiset of intervals

$$\mathrm{Bar}_k(X) = \{(r_{i_1}, r_{j_1}), (r_{i_2}, r_{j_2}), \dots\},$$

where the (r_i, r_j) are the indices appearing in the decomposition

$$\mathcal{M}_k = \mathbb{I}_{r_{i_1}, r_{j_1}} \oplus \mathbb{I}_{r_{i_2}, r_{j_2}} \oplus \dots$$

. The r_i and r_j birth times and death times, respectively.

Remark

Elements of $\mathrm{Bar}_k(X)$ that are "long" (with respect to some statistical significance level) represent statistically significant homological features of X.

Remark

Elements of $\mathrm{Bar}_k(X)$ that are "long" (with respect to some statistical significance level) represent statistically significant homological features of X.

Representation Theory of Barcodes

The category of persistence modules over a commutative ring R is equivalent to the category of finitely generated graded modules over $R[t]$.

In particular, if R is a field, then any barcode is encoded as an endomorphism of a real vector space.

Persistence in Practice

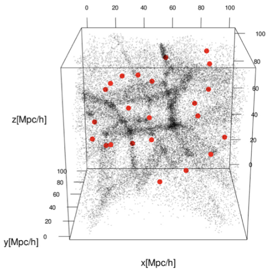
Persistent homology is very useful for analyzing the shape of noisy data sets.

Cosmological data

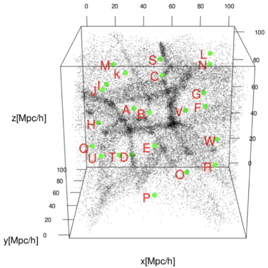
The problem of detecting

- voids (H_2 generators),
- loops (H_1 generators),
- clusters (H_0 generators)

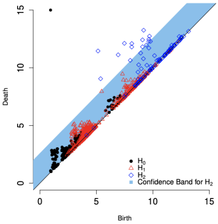
in points of mass sampled from the "cosmic web."



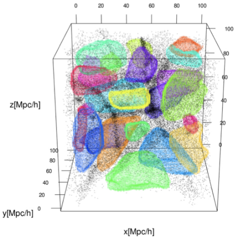
(a) Voronoi foam data



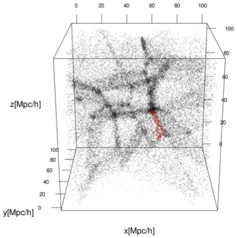
(b) Volume centers



(c) Persistence diagram



(d) Example voids



(e) Example filament loop

Table of Contents

Proposition (from before)

The multiplicity of the k^{th} 0-eigenvalue of the Hodge Laplacian on a smooth manifold counts generators of the k^{th} simplicial homology.

Now, let

$$X_{r_0} \hookrightarrow X_{r_1} \hookrightarrow X_{r_2} \hookrightarrow \dots$$

be an increasing filtration of (almost) manifolds.

We can develop an analogous theory of persistent Laplacians for this filtration. In particular, the Proposition implies that the **persistent Laplacian is strictly stronger than the persistent homology.**

Persistent Hodge theory

As discussed earlier, the spectrum of the Laplacian encodes important geometric information, but does not uniquely determine the geometry. That is, we cannot hear the shape of a drum.

A reasonable step, then, is to analyze the evolution of the spectra over a filtration of manifolds, to understand which geometric features are informed by the Laplacian spectra.

Hearing shape.

Let E_1 and E_2 be plane regions bounded by curves Γ_1 and Γ_2, with associated eigenvalues and eigenfunctions denoted by $\{\lambda_i, \psi_i\}_{i\in I}$ and $\{\lambda_i', \psi_i'\}_{i\in I}$, respectively. Furthermore, assume that $\lambda_i = \lambda_i'$ for all i.

Hearing shape.

Let E_1 and E_2 be plane regions bounded by curves Γ_1 and Γ_2, with associated eigenvalues and eigenfunctions denoted by $\{\lambda_i, \psi_i\}_{i\in I}$ and $\{\lambda_i', \psi_i'\}_{i\in I}$, respectively. Furthermore, assume that $\lambda_i = \lambda_i'$ for all i.

Question

Are E_1 and E_2 congruent (*i.e.* isometric)?

Hearing shape.

Let E_1 and E_2 be plane regions bounded by curves Γ_1 and Γ_2, with associated eigenvalues and eigenfunctions denoted by $\{\lambda_i, \psi_i\}_{i\in I}$ and $\{\lambda_i', \psi_i'\}_{i\in I}$, respectively. Furthermore, assume that $\lambda_i = \lambda_i'$ for all i.

Question

Are E_1 and E_2 congruent (*i.e.* isometric)?

Answer

No. There are several examples of E_1, E_2 that are isospectral but not isometric. Thus we can not deduce the shape of a drum based on its harmonic overtones.

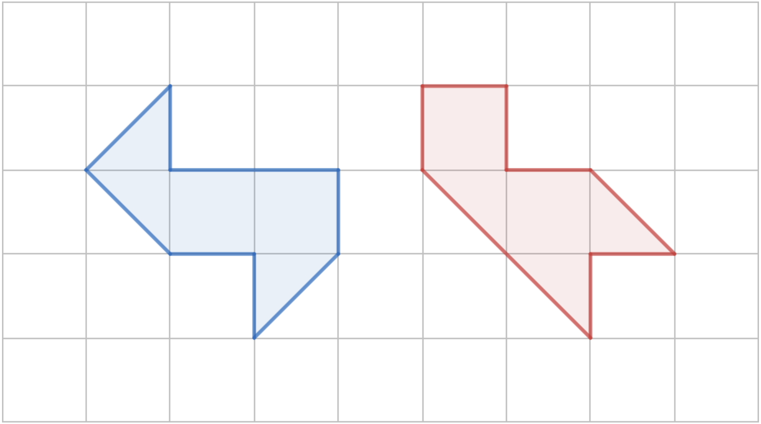


Figure 6: Two isospectral but not isometric plane regions.

Interpreting Laplacian spectra.

If they can't be used to infer shape, what are Laplacian spectra useful for?

1. Determining area.

Interpreting Laplacian spectra.

If they can't be used to infer shape, what are Laplacian spectra useful for?

1. Determining area.
2. Graph theory.

Interpreting Laplacian spectra.

If they can't be used to infer shape, what are Laplacian spectra useful for?

1. Determining area.
2. Graph theory.
3. Dimensionality reduction.

Interpreting Laplacian spectra.

If they can't be used to infer shape, what are Laplacian spectra useful for?

1. Determining area.
2. Graph theory.
3. Dimensionality reduction.
4. Signal processing.

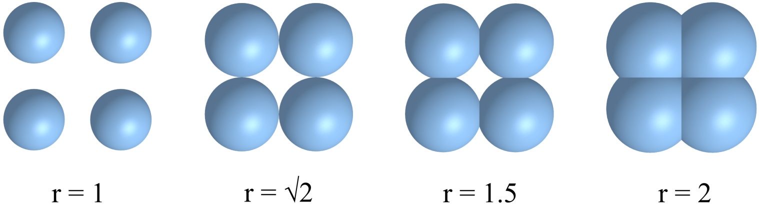


Figure 7: A filtration of 2-manifolds. Persistent homology does not capture *e.g.* the change in homotopy type from $r = \sqrt{2}$ to $r = \sqrt{2} + \epsilon$. That is, persistent Hodge theory has a richer structure which captures geometric changes in the absence of (co)homological changes.

It is possible to define an analog of the Hodge Laplacian that captures the evolution of spectra in a filtration of manifolds.

Proposition

There exists a differential operator, the b-fold k^{th} Hodge Laplacian on the manifold X_a

$$\triangle_k^{a,b} : \Omega^k(X_a) \to \Omega^k(X_a)$$

whose kernel is precisely the harmonic k-forms on X_a whose images under $\iota : X_a \hookrightarrow X_b$ are also harmonic.

It is possible to define an analog of the Hodge Laplacian that captures the evolution of spectra in a filtration of manifolds.

Proposition

There exists a differential operator, the b-fold k^{th} Hodge Laplacian on the manifold X_a

$$\triangle_k^{a,b} : \Omega^k(X_a) \to \Omega^k(X_a)$$

whose kernel is precisely the harmonic k-forms on X_a whose images under $\iota : X_a \hookrightarrow X_b$ are also harmonic.

Remark

Furthermore, the spectrum of $\triangle_k^{a,b}$ is precisely the spectral elements that persist from a to b in the filtration.

Properties

Let $\lambda_k^{l,j}$ denote the j^{th} least eigenvalue of $\triangle_k$ on $\Omega(X_l)$. Then, from Hodge theory

- The birth of a zero eigenvalue corresponds to the birth of a topological feature.
- The birth of a non-zero eigenvalue corresponds to the death of a topological feature.

Remark

Hence, analysis of the Hodge Laplacians recovers all of the information given by the persistent homology.

Properties

Let $\lambda_k^{l,j}$ denote the j^{th} least eigenvalue of $\triangle_k$ on $\Omega(X_l)$. Then, from Hodge theory

- The birth of a zero eigenvalue corresponds to the birth of a topological feature.
- The birth of a non-zero eigenvalue corresponds to the death of a topological feature.

Remark

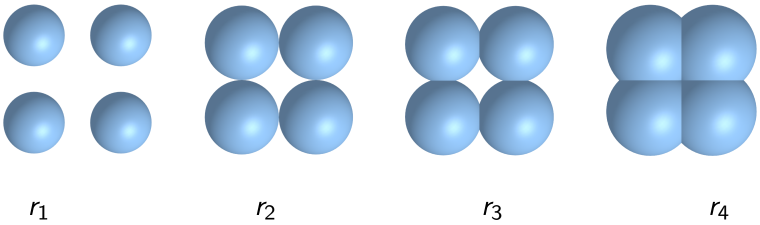
Hence, analysis of the Hodge Laplacians recovers all of the information given by the persistent homology. This is a bit of a lie.

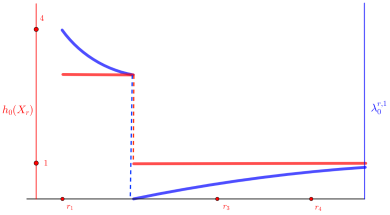
Together, these ideas suggest that

The evolution of the $\lambda_k^{l,1}$ may indicate upcoming (with respect to the filtration) of births and deaths of topological features.

This might be realized as the derivative(s) (with respect to the filtration) of the spectra $\triangle_k^{a,b}$.

However, in practice this requires our filtration to be indexed over a continuous moduli space B (*e.g.* $\mathbb{R}$), and it remains to be shown that $\lambda_k^{l,j}$ varies differentiably with respect to $j \in B$.


r_1
r_2
r_3
r_4


$h_0(X_r)$
4
1
r_1
r_3
r_4
$\lambda_0^{r,1}$

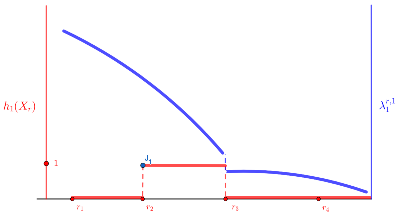

$h_1(X_r)$
1
r_1
r_2
r_3
r_4
$\lambda_1^{r,1}$

Table of Contents

Summary

Persistent homology captures useful topological information about filtrations of manifolds, particularly noisy point cloud data.

However, it gives strictly less information than analyzing the evolution of Hodge Laplacian spectra over the filtration.

In the same way that classical Hodge theory compares topological and geometric data, I hope to compare topological and geometric features of data sets using persistent Hodge theory.

More importantly, it provides a finer way of measuring geometric progression of evolving data sets along a time parameter.

Another potential direction is to combine the techniques from persistent homology and persistent Hodge theory. That is, to compute the progression of Betti numbers *and* eigenvalues jointly over a 2-dimensional parameter space.

There is a robust theory for computing persistent homology varying over non partially ordered parameter spaces. There are few obstructions extending this framework to persistent Hodge theory.

[1] S. J. Chapman. "Drums That Sound the Same". In: *The American Mathematical Monthly* 102.2 (Feb. 1995), p. 124. ISSN: 00029890. DOI: 10.2307/2975346. URL: https://www.jstor.org/stable/2975346?origin=crossref (visited on 09/12/2023).

[2] Justin Curry. "Topological Data Analysis and Cosheaves". In: (2014). Publisher: arXiv Version Number: 2. DOI: 10.48550/ARXIV.1411.0613. URL: https://arxiv.org/abs/1411.0613 (visited on 09/14/2023).

[3] Yongyuth Eiamsa-ard et al. "ASEAN Studies: The Revival and Development of Lao Classical Music.". In: *Turkish Online Journal of Qualitative Inquiry* 12.7 (2021), pp. 2323–2331. ISSN: 13096591. URL: https://search.ebscohost.com/login.aspx?direct=true&db=eft&AN=161811635&site=eds-live.

[4] Carolyn Gordon, David L. Webb, and Scott Wolpert. "One cannot hear the shape of a drum". In: (1992). Publisher: arXiv Version Number: 1. DOI: 10.48550/ARXIV.MATH/9207215. URL: https://arxiv.org/abs/math/9207215 (visited on 09/14/2023).

[5] Kōichi Iwabuchi, Eva Tsai, and Chris Berry, eds. *Routledge handbook of East Asian popular culture*. OCLC: 965196416. Abingdon, Oxon: Routledge, 2017. ISBN: 978-1-317-28501-4.

[6] Mark Kac. “Can One Hear the Shape of a Drum?” In: *The American Mathematical Monthly* 73.4 (Apr. 1966), p. 1. ISSN: 00029890. DOI: 10.2307/2313748. URL: https://www.jstor.org/stable/2313748?origin=crossref (visited on 09/12/2023).

[7] Terry Miller and Jarernchai Chonpairot. “The Musical Traditions of Northeast Thailand”. In: *The Journal of the Siam Society* 67.1 (1979). URL: https://thesiamsociety.org/wp-content/uploads/1979/03/JSS_067_1b_MillerJarernchaiChonpairot_MusicalTraditionsOfNortheastThailand.pdf.

[8] James Norman Mosel. “Trends and structure in contemporary Thai poetry, with translations and bibliography”. In: ().

[9] Boike Rehbein. *Globalization, culture and society in laos*. OCLC: 671531702. Place of publication not identified: Routledge, 2010. ISBN: 978-0-415-59218-5.

[10] Emily Ribando-Gros et al. *Graph and Hodge Laplacians: Similarity and Difference*. Apr. 26, 2022. arXiv: 2204.12218[math]. URL: http://arxiv.org/abs/2204.12218 (visited on 09/14/2023).

[11] Thanyalak Ubonlert, Thinakorn Atapaiboon, and Chaloey Pumipuntu. "Guideline for the Development and Management of Music and Performing Arts in Khwaeng Champasak, Lao PDR". In: *Asian Culture and History* 7.1 (Sept. 24, 2014), p126. ISSN: 1916-9663, 1916-9655. DOI: 10.5539/ach.v7n1p126. URL: http://www.ccsenet.org/journal/index.php/ach/article/view/38082 (visited on 05/10/2023).

[12] Claire Voisin. *Hodge Theory and Complex Algebraic Geometry I*. Trans. by Leila Schneps. 1st ed. Cambridge University Press, Dec. 5, 2002. ISBN: 978-0-521-80260-4 978-0-521-71801-1 978-0-511-61534-4. DOI: 10.1017/CBO9780511615344. URL: https://www.cambridge.org/core/product/identifier/9780511615344/type/book (visited on 09/14/2023).

[13] Xin Xu et al. "Finding cosmic voids and filament loops using topological data analysis". In: (2018). Publisher: arXiv Version Number: 2. DOI: 10.48550/ARXIV.1811.08450. URL: https://arxiv.org/abs/1811.08450 (visited on 09/14/2023).