Persistent homology, metrics on diagrams and metric space valued functions

Peter Bubenik

Department of Mathematics
Cleveland State University
p.bubenik@csuohio.edu
http://academic.csuohio.edu/bubenik_p/

July 25, 2013
joint work with Vin de Silva and Jonathan A. Scott

funded by AFOSR FA9550-13-1-0115
Overview

Data → Filtered simplicial complex → Homology → Persistence module → Barcode

Peter Bubenik Persistent homology and metrics on diagrams
Overview

Data \rightarrow \text{Filtered simplicial complex} \xrightarrow{\text{Homology}} \text{Persistence module} \rightarrow \text{Barcode}

Metric

Peter Bubenik
Persistent homology and metrics on diagrams
Overview

Data → Filtered simplicial complex → Homology → Persistence module → Barcode

Metric → Metric → Metric → Metric → Metric
Overview

Data → Filtered simplicial complex → Homology → Persistence module → Barcode

Metric → Metric → Metric → Metric

Peter Bubenik
Persistent homology and metrics on diagrams
Overview

Data \rightarrow \text{Filtered simplicial complex} \rightarrow \text{Homology} \rightarrow \text{Persistence module} \rightarrow \text{Barcode}

Metric \rightarrow \text{Metric} \rightarrow \text{Metric} \rightarrow \text{Metric}

? \rightarrow \text{Diagram} \rightarrow \text{Generalized persistence module} \rightarrow ?

Persistent homology and metrics on diagrams
Overview

Data → Filtered simplicial complex → Homology → Persistence module → Barcode

Metric → Metric → Metric → Metric

? → Diagram → Generalized persistence module → ?

Structure → Metric
Overview

Data → Filtered simplicial complex → Homology → Persistence module → Barcode

Function to a metric space → Diagram → Generalized persistence module → ?

Metric → Metric → Metric → Metric → Metric

Metric → Structure → Metric

Peter Bubenik
Persistent homology and metrics on diagrams
Topological data analysis

From data to topology:
1. Start with a finite set of points in some metric space.
2. Apply a geometric construction (e.g. Čech, Rips) to obtain a nested sequence of simplicial complexes.
Topological data analysis

From data to topology:

1. Start with a finite set of points in some metric space.
2. Apply a geometric construction (e.g. Čech, Rips) to obtain a nested sequence of simplicial complexes.
Topological data analysis

From data to topology:
1. Start with a finite set of points in some metric space.
2. Apply a geometric construction (e.g. Čech, Rips) to obtain a nested sequence of simplicial complexes.
Topological data analysis

From data to topology:

1. Start with a finite set of points in some metric space.
2. Apply a geometric construction (e.g. Čech, Rips) to obtain a nested sequence of simplicial complexes.
Topological data analysis

From data to topology:

1. Start with a finite set of points in some metric space.
2. Apply a geometric construction (e.g. Čech, Rips) to obtain a nested sequence of simplicial complexes.
Topological data analysis

From data to topology:

1. Start with a finite set of points in some metric space.
2. Apply a geometric construction (e.g., Čech, Rips) to obtain a nested sequence of simplicial complexes.
Topological data analysis

From data to topology:

1. Start with a finite set of points in some metric space.
2. Apply a geometric construction (e.g. Čech, Rips) to obtain a nested sequence of simplicial complexes.
Topological data analysis

From data to topology:

1. Start with a finite set of points in some metric space.
2. Apply a geometric construction (e.g. Čech, Rips) to obtain a nested sequence of simplicial complexes.
Topological data analysis

From data to topology:

1. Start with a finite set of points in some metric space.
2. Apply a geometric construction (e.g. Čech, Rips) to obtain a nested sequence of simplicial complexes.
Topological data analysis

From data to topology:

1. Start with a finite set of points in some metric space.
2. Apply a geometric construction (e.g. Čech, Rips) to obtain a nested sequence of simplicial complexes.
Topological data analysis

From data to topology:

1. Start with a finite set of points in some metric space.
2. Apply a geometric construction (e.g. Čech, Rips) to obtain a nested sequence of simplicial complexes.
Topological data analysis

From data to topology:

1. Start with a finite set of points in some metric space.
2. Apply a geometric construction (e.g. Čech, Rips) to obtain a nested sequence of simplicial complexes.
Topological data analysis

From data to topology:

1. Start with a finite set of points in some metric space.
2. Apply a geometric construction (e.g. Čech, Rips) to obtain a nested sequence of simplicial complexes.
Topological data analysis

From data to topology:

1. Start with a finite set of points in some metric space.
2. Apply a geometric construction (e.g. Čech, Rips) to obtain a nested sequence of simplicial complexes.
Persistent homology

We have a nested sequence of simplicial complexes,

\[ K_0 \rightarrow K_1 \rightarrow \cdots \rightarrow K_n. \]  

(\(*\) )

Apply simplicial homology,

\[ H(K_0) \rightarrow H(K_1) \rightarrow \cdots \rightarrow H(K_n). \]  

(\(H\*) )
Persistent homology

We have a nested sequence of simplicial complexes,

$$K_0 \rightarrow K_1 \rightarrow \cdots \rightarrow K_n.$$  \hfill (\ast)

Apply simplicial homology,

$$H(K_0) \rightarrow H(K_1) \rightarrow \cdots \rightarrow H(K_n).$$  \hfill (H\ast)

The shape of these diagrams is given by the category $n$,

$$0 \rightarrow 1 \rightarrow \cdots \rightarrow n.$$

Then (\ast) is equivalent to $n \xrightarrow{K} \text{Simp}$, and (H\ast) is equivalent to $n \xrightarrow{K} \text{Simp} \xrightarrow{H} \text{Vect}_F$. 

Peter Bubenik
Persistent homology and metrics on diagrams
Another paradigm:

1. Start with a function $f : X \rightarrow \mathbb{R}$.
2. For each $a \in \mathbb{R}$, consider $f^{-1}((\mathbb{R}, \leq))$.
3. This gives us a diagram $F : (\mathbb{R}, \leq) \rightarrow \text{Top}$.

Composing with singular homology we have,

$$(\mathbb{R}, \leq) \xrightarrow{F} \text{Top} \xrightarrow{H} \text{Vect}_F.$$
Multidimensional persistent homology

1. Start with a function $f : X \rightarrow \mathbb{R}^n$.

2. For each $a \in \mathbb{R}^n$, consider $f^{-1}(\mathbb{R}^n_{\leq a})$.
   - This gives us a diagram $F : (\mathbb{R}^n, \leq) \rightarrow \text{Top}$.

3. Composing with singular homology we have,

\[
(\mathbb{R}^n, \leq) \xrightarrow{F} \text{Top} \xrightarrow{H} \text{Vect}_F.
\]
Levelset persistent homology

1. Start with a function \( f : X \to \mathbb{R} \).
2. For each interval \( I \subseteq \mathbb{R} \), consider \( f^{-1}(I) \).
   - This gives us a diagram \( F : \text{Intervals} \to \text{Top} \).
3. Composing with singular homology we have,
   \[
   \text{Intervals} \xrightarrow{F} \text{Top} \xrightarrow{H} \text{Vect}_F.
   \]
Angle-valued persistent homology

1. Start with a function \( f : X \rightarrow S^1 \).

2. For each arc \( A \subseteq S^1 \), consider \( f^{-1}(A) \).
   - This gives us a diagram \( F : \text{Arcs} \rightarrow \text{Top} \).

3. Composing with singular homology we have,

\[
\text{Arcs} \xrightarrow{F} \text{Top} \xrightarrow{H} \text{Vect}_F.
\]
We will use category theory to give a unified treatment of each of the above flavors of persistent homology.

Why?

- Give simpler, common proofs to some basic persistence results.
- Remove assumptions.
- Apply persistence to functions, $f : X \to (M, d)$.
- Allow homology to be replaced with other functors.
- Provide a framework for new applications.

Specific goal:

- Interpret and prove stability in this setting.
Termiology

In this talk a metric will be allowed to

- have \( d(x, y) = \infty \) for \( x \neq y \), and
- have \( d(x, y) = 0 \) for \( x \neq y \).

That is, it is an extended pseudometric.

Example: The Hausdorff distance on the set of all subspaces of \( \mathbb{R} \).
Unified framework

Generalized persistence module,

\[ \mathcal{P} \xrightarrow{F} \mathcal{C} \xrightarrow{H} \mathcal{A}. \]

Here,

- The indexing category \( \mathcal{P} \) is a poset together with some notion of distance;
- \( \mathcal{C} \) is some category;
- \( \mathcal{A} \) is some abelian category (e.g. \( \text{Vect}_F \), \( \text{R-mod} \));
- \( F \) and \( H \) are arbitrary functors.
Main results

Theorem (Interleaving distance)

There is a distance function \(d(F, G)\) between diagrams \(F, G : \mathbf{P} \to \mathbf{C}\). This ‘interleaving distance’ is a metric.

Theorem (Stability of interleaving distance)

Let \(F, G : \mathbf{P} \to \mathbf{C}\) and \(H : \mathbf{C} \to \mathbf{A}\). Then,

\[d(H \circ F, H \circ G) \leq d(F, G).\]
Inverse images of metric space valued functions

Start with \( f : X \rightarrow (M, d_M) \).
Let \( P \) be a poset of subsets of \((M, d_M)\).
Define
\[
F : P \rightarrow \text{Top} \\
U \mapsto f^{-1}(U)
\]
Inverse images of metric space valued functions

Start with $f : X \rightarrow (M, d_M)$.
Let $\mathbf{P}$ be a poset of subsets of $(M, d_M)$.
Define

$$F : \mathbf{P} \rightarrow \text{Top}$$

$$U \mapsto f^{-1}(U)$$

**Theorem (Inverse-image stability)**

Let $F, G : \mathbf{P} \rightarrow \text{Top}$ be given by inverse images of $f, g : X \rightarrow (M, d_M)$.

$$d(F, G) \leq d_{\infty}(f, g) := \sup_{x \in X} d_M(f(x), g(x)).$$

**Corollary (Stability of generalized persistence modules)**

Let $H : \text{Top} \rightarrow A$. Then

$$d(HF, HG) \leq d_{\infty}(f, g).$$
Examples

<table>
<thead>
<tr>
<th>persistence</th>
<th>( M )</th>
<th>( P )</th>
</tr>
</thead>
<tbody>
<tr>
<td>ordinary</td>
<td>( \mathbb{R} )</td>
<td>( { (-\infty, a] \mid a \in \mathbb{R} } )</td>
</tr>
<tr>
<td>multidimensional</td>
<td>( \mathbb{R}^n )</td>
<td>( { \mathbb{R}^n_\leq a \mid a \in \mathbb{R}^n } )</td>
</tr>
<tr>
<td>levelset</td>
<td>( \mathbb{R} )</td>
<td>( { \text{intervals in } \mathbb{R} } )</td>
</tr>
<tr>
<td>angle-valued</td>
<td>( S^1 )</td>
<td>( { \text{arcs in } S^1 } )</td>
</tr>
</tbody>
</table>

For each of these examples,

- \( P \) is a poset under inclusion;
- for \( f : X \to M \), \( F : P \to \text{Top} \) is given by inverse images of \( f \);
- for \( f, g : X \to M \) and \( H : \text{Top} \to A \), \( d(HF, HG) \leq d_\infty(f, g) \).
A **natural transformation** is a map of diagrams. For \( V, W : n \rightarrow \text{Vect}_F \) it is a commutative diagram,

\[
\begin{align*}
V_0 & \rightarrow V_1 \rightarrow \cdots \rightarrow V_n \\
\varphi_0 & \downarrow \varphi_1 & & \ldots & & \varphi_n \\
W_0 & \rightarrow W_1 \rightarrow \cdots \rightarrow W_n
\end{align*}
\]

For \( F, G : P \rightarrow D \), for all \( x \leq y \) there is a commuting diagram,

\[
\begin{align*}
F(x) & \xrightarrow{F(x \leq y)} F(y) \\
\varphi_x & \downarrow \varphi_y \\
G(x) & \xrightarrow{G(x \leq y)} G(y)
\end{align*}
\]
Comparing diagrams

We denote a natural transformation by $\varphi : F \Rightarrow G$.

Two diagrams $F$, $G$ are isomorphic if we have $\varphi : F \Rightarrow G$ and $\psi : G \Rightarrow F$ such that $\psi \circ \varphi = \text{Id}$ and $\varphi \circ \psi = \text{Id}$.

What if $F$ and $G$ are not isomorphic?
Comparing diagrams

We denote a natural transformation by $\varphi : F \Rightarrow G$.

Two diagrams $F$, $G$ are **isomorphic** if we have $\varphi : F \Rightarrow G$ and $\psi : G \Rightarrow F$ such that $\psi \circ \varphi = \text{Id}$ and $\varphi \circ \psi = \text{Id}$.

What if $F$ and $G$ are not isomorphic?

We would like to be able to quantify how far $F$ and $G$ are from being isomorphic.

- We will define **translations** on $\mathbf{P}$, and use these
- to define **interleavings** between diagrams.
- Then a metric on $\mathbf{P}$ will give us the **interleaving distance**.
Translations

**Definition**

A **translation** is given by $\Gamma : P \rightarrow P$ such that $x \leq \Gamma(x)$ for all $x$.

Example:

$$\Gamma : (\mathbb{R}, \leq) \rightarrow (\mathbb{R}, \leq)$$

$$a \mapsto a + \varepsilon$$

- The identity is a translation.
- The composition of translations is a translation.
Interleaving

**Definition**

F and G are \((\Gamma, K)\)-interleaved if there exist \(\varphi, \psi\),

\[
\begin{array}{c}
\text{P} \\ \Gamma \\ F \\
\varphi \\
\text{C} \\
\hline
\text{P} \\ K \\ G \\ \psi \\
\text{C} \\
\hline
\text{P} \\ F \\
\end{array}
\]

such that

\[
\psi \varphi = FK\Gamma \text{ and } \varphi \psi = G\Gamma K.
\]
Interleaving of \((\mathbb{R}, \leq)\)-indexed diagrams

\[ \Gamma = K : a \mapsto a + \varepsilon \]
\[ \forall a \leq b, \]
\[ F(a) \rightarrow F(b) \]
\[ G(a + \varepsilon) \rightarrow G(b + \varepsilon) \]
\[ F(a + \varepsilon) \rightarrow F(b + \varepsilon) \]
\[ G(a + \varepsilon) \rightarrow G(b + \varepsilon) \]
\[ G(a) \rightarrow G(b) \]
\[ \forall a, \]
\[ F(a) \rightarrow F(a + 2\varepsilon) \]
\[ G(a + \varepsilon) \]
\[ G(a + 2\varepsilon) \]
\[ F(a + \varepsilon) \rightarrow F(a + 2\varepsilon) \]
\[ G(a) \rightarrow G(a + 2\varepsilon) \]
Now assume that $P$ has a metric $d$.

An $\varepsilon$-translation is a translation $\Gamma : P \rightarrow P$ such that $d(x, \Gamma(x)) \leq \varepsilon$ for all $x$.

**Definition**

$$d(F, G) = \inf(\varepsilon \mid F, G \text{ interleaved by } \varepsilon\text{-translations})$$

**Theorem (Interleaving distance)**

*This interleaving distance is metric.*
Given \((M, d_M)\), let \(P\) be a subset of \(\mathcal{P}(M)\) with partial order given by inclusion, and Hausdorff distance.

<table>
<thead>
<tr>
<th>persistence</th>
<th>(M)</th>
<th>(P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ordinary</td>
<td>(\mathbb{R})</td>
<td>({(-\infty, a] \mid a \in \mathbb{R}})</td>
</tr>
<tr>
<td>multidimensional</td>
<td>(\mathbb{R}^n)</td>
<td>(\mathbb{R}^n_{\leq a} \mid a \in \mathbb{R}^n)</td>
</tr>
<tr>
<td>levelset</td>
<td>(\mathbb{R})</td>
<td>{intervals in (\mathbb{R}}}</td>
</tr>
<tr>
<td>angle-valued</td>
<td>(S^1)</td>
<td>{arcs in (S^1}}</td>
</tr>
</tbody>
</table>

- Particular \(\varepsilon\)-translations, \(\Gamma_\varepsilon\), are given by thickening by \(\varepsilon\).
- Note that each poset \(P\) is closed under these \(\Gamma_\varepsilon\).
Using functoriality

**Theorem (Stability of interleaving distance)**

Let $F, G : \mathbf{P} \to \mathbf{C}$ and $H : \mathbf{C} \to \mathbf{A}$. Then,

$$d(H \circ F, H \circ G) \leq d(F, G).$$

**Proof.**

\[
\begin{array}{ccc}
\mathbf{P} & \xrightarrow{\Gamma} & \mathbf{P} & \xrightarrow{K} & \mathbf{P} \\
\downarrow F & & \downarrow G & & \downarrow F \\
\mathbf{C} & \xrightarrow{\varphi} & \mathbf{C} & \xrightarrow{\psi} & \mathbf{C} \\
\downarrow H & = & \downarrow H & = & \downarrow H \\
\mathbf{A} & = & \mathbf{A} & = & \mathbf{A}
\end{array}
\]
Inverse-image stability

**Theorem (Inverse-image stability)**

Let $F, G : P \to \textbf{Top}$ correspond to $f, g : X \to (M, d_M)$. Assume $P$ closed under $\Gamma_\varepsilon$ for all $\varepsilon$.

\[ d(F, G) \leq d_\infty(f, g) := \sup_{x \in X} d_M(f(x), g(x)). \]

**Proof.**

Let $\varepsilon = d_\infty(f, g)$. $F, G$ are $\varepsilon$-interleaved:

\[ F(S) = f^{-1}(S) \subseteq g^{-1}(\Gamma_\varepsilon(S)) = G\Gamma_\varepsilon(S). \]

Thus,

\[ d(F, G) \leq \varepsilon = d_\infty(f, g). \]
Algebraic structure

Until this point we have only imposed structure on the indexing category: a partial order and a metric.

To compute we need some algebraic structure in our target category. For example, $\text{Vect}_{\mathbb{F}_2}$, $\text{Vect}_{\mathbb{F}}$, or $\text{Ab}$. 
Algebraic structure

Until this point we have only imposed structure on the indexing category: a partial order and a metric.

To compute we need some algebraic structure in our target category. For example, $\text{Vect}_{\mathbb{F}_2}$, $\text{Vect}_{\mathbb{F}}$, or $\text{Ab}$.

We will assume the target category, $\mathcal{A}$, is abelian.

This also includes $\mathcal{R}$-mod and sheaves of abelian groups on $X$.

**Definition**

A category, $\mathcal{A}$, is **abelian** if

- each hom-set is an abelian group;
- all finite direct sums exist; and
- all morphisms have kernels and cokernels.
Kernel, image and cokernel persistence

Given

- $X \subseteq Y \in \textbf{Top}$ and $g : Y \to (M, d_M)$;
- $\mathcal{P}$ is a poset of subsets of $M$ closed under $\Gamma_\varepsilon$, and
- $H : \textbf{Top} \to \mathbf{A}$.

Let $f : X \hookrightarrow Y \overset{g}{\to} (M, d_M)$.
Let $F, G : \mathcal{P} \to \textbf{Top}$ be given by inverse images of $f, g$.

Since $f^{-1}(U) \subseteq g^{-1}(U)$, $F \hookrightarrow G$, and $HF \overset{\alpha}{\to} HG \in \mathbf{A}^\mathcal{P}$.

Since $\mathbf{A}$ is abelian, so is $\mathbf{A}^\mathcal{P}$.
So the kernel, image and cokernel of $\alpha$ exist.
Stability for kernel, image and cokernel persistence

Given \( X \subseteq Y \in \textbf{Top} \) and \( g, g' : Y \rightarrow (M, d_M) \).

Construct \( \alpha : HF \rightarrow HG \) and \( \alpha' : HF' \rightarrow HG' \) as above.

**Theorem (Stability of ker/im/coker persistence)**

\[
\begin{align*}
    d(\ker(\alpha), \ker(\alpha')) & \leq d_\infty(g, g') \\
    d(\text{im}(\alpha), \text{im}(\alpha')) & \leq d_\infty(g, g') \\
    d(\text{coker}(\alpha), \text{coker}(\alpha')) & \leq d_\infty(g, g')
\end{align*}
\]
Monoid of translations

Recall def of translation: $\Gamma : P \to P$ such that $x \leq \Gamma(x)$ for all $x$. That is, $\Gamma$ is an endofunctor of $P$ together with $\text{Id} \Rightarrow \Gamma$. 
Monoid of translations

Recall def of translation: \( \Gamma : P \rightarrow P \) such that \( x \leq \Gamma(x) \) for all \( x \).

That is, \( \Gamma \) is an endofunctor of \( P \) together with \( \text{Id} \Rightarrow \Gamma \).

\( \text{End}_*(P) \) contain \( \text{Id} \) and are closed under composition.
So it is a monoid.

Two ways of defining interleaving distance:

\[
\text{End}_*(P) \xrightarrow{\rho} ([0, \infty], \leq, +, 0)
\]

\[\rho : \Gamma \mapsto \sup(d(x, \Gamma(x)))\]

\[\iota : \varepsilon \mapsto \Gamma_\varepsilon\]
Main theorem

Our main results

- the existence of an interleaving distance, and
- the stability of this interleaving distance,

can be summarized as follows.

**Theorem**

*Given a poset $P$ with a metric, the interleaving distance gives a functor*

$$\text{Cat}(P, -) : \text{Cat} \rightarrow \text{Metric}.$$