Scalar field analysis with aberrant noise

Mickaël Buchet

joint work with F. Chazal, T. Dey, F. Fan, S. Oudot and Y. Wang
What is scalar field analysis?

How many peaks do you see?
What is scalar field analysis?

How many peaks do you see?
What is scalar field analysis?

How many peaks do you see?
What is scalar field analysis?

Persistence diagram
Persistence diagram
What is scalar field analysis?

Persistence diagram
Persistence diagram
What is scalar field analysis?

Persistence diagram
What is scalar field analysis?

Persistence diagram
What is scalar field analysis?

Persistence diagram
Persistence diagram
What is scalar field analysis?

Persistence diagram
Persistence diagram
What is scalar field analysis?

Persistence diagram
Comparison between persistence diagrams
Comparison between persistence diagrams
Comparison between persistence diagrams

What is scalar field analysis?

Bottleneck distance:

$$d_B(D, E) = \inf_{b \in B} \max_{x \in D} ||x - b(x)||_\infty$$
What is scalar field analysis?

Comparison between persistence diagrams

Bottleneck distance:

\[ d_B(D, E) = \inf_{b \in B} \max_{x \in D} \| x - b(x) \|_\infty \]
Higher dimensions
What is scalar field analysis?

Applicative setting

The *ground truth*:

- a sub-manifold $M$ of $\mathbb{R}^d$
- a $c$-Lipschitz function $f : M \mapsto \mathbb{R}$
Applicative setting

The *ground truth*:

- a sub-manifold $M$ of $\mathbb{R}^d$
- a $c$-Lipschitz function $f : M \mapsto \mathbb{R}$

What we really know:

- a set of points $P \in \mathbb{R}^d$
- a function $\tilde{f} : P \mapsto \mathbb{R}$
Applicative setting

The ground truth:
- a sub-manifold $M$ of $\mathbb{R}^d$
- a $c$-Lipschitz function $f : M \mapsto \mathbb{R}$

What we really know:
- a set of points $P \in \mathbb{R}^d$
- a function $\tilde{f} : P \mapsto \mathbb{R}$

Approximate the persistence diagram $D$ of $f$ by a diagram $\hat{D}$
Previous work

From Chazal, Guibas, Oudot and Skraba (DCG’11)
Under some conditions on $\epsilon$ and the geometry of $M$,

Theorem

*If $P$ is an $\epsilon$ Riemannian sample of $M$ and $\|f|_P - \tilde{f}\|_{\infty} \leq \xi$, then:*
Previous work

From Chazal, Guibas, Oudot and Skraba (DCG’11)
Under some conditions on $\epsilon$ and the geometry of $M$,

**Theorem**

*If $P$ is an $\epsilon$ Riemannian sample of $M$ and $\|f|_P - \tilde{f}\|_{\infty} \leq \xi$, then:*

$$d_B(D, \hat{D}) \leq 4c\epsilon + \xi$$
Previous work

From Chazal, Guibas, Oudot and Skraba (DCG’11)
Under some conditions on $\epsilon$ and the geometry of $M$,

**Theorem**
*If $P$ is an $\epsilon$ Riemannian sample of $M$ and $\|f_P - \tilde{f}\|_{\infty} \leq \xi$, then:*

$$d_B(D, \hat{D}) \leq 4c\epsilon + \xi$$

**Theorem**
*If $P$ is an $\epsilon$ Riemannian sample of $M$ and the pairwise distances between points of $P$ are known with precision $\nu$, then:*

...
What is scalar field analysis?

Previous work

From Chazal, Guibas, Oudot and Skraba (DCG’11)
Under some conditions on $\epsilon$ and the geometry of $M$,

**Theorem**

*If $P$ is an $\epsilon$ Riemannian sample of $M$ and $||f|_P - \tilde{f}||_{\infty} \leq \xi$, then:*

$$d_B(D, \hat{D}) \leq 4c\epsilon + \xi$$

**Theorem**

*If $P$ is an $\epsilon$ Riemannian sample of $M$ and the pairwise distances between points of $P$ are known with precision $\nu$, then:*

$$d_B(D, \hat{D}) \leq (4\epsilon + 2\nu)c$$
What is scalar field analysis?

A bad example
What is scalar field analysis?

A bad example
What is scalar field analysis?

A bad example

![Graph](image.png)
Filtered simplicial complex

Classical algorithms to compute a persistence diagram work on a filtered simplicial complex.
Filtered simplicial complex

Classical algorithms to compute a persistence diagram work on a filtered simplicial complex.

A simplicial complex is a set $X$ of simplices such that for any simplex $\sigma \in X$, all facets of $\sigma$ are also in $X$. 
Filtered simplicial complex

Classical algorithms to compute a persistence diagram work on a filtered simplicial complex.

A simplicial complex is a set $X$ of simplices such that for any simplex $\sigma \in X$, all facets of $\sigma$ are also in $X$.

We say that $X$ is filtered when there exists a family of simplicial complexes $\{X_\alpha\}_{\alpha \in \mathbb{R}}$ such that for all $\alpha < \beta$, $X_\alpha \subset X_\beta \subset X$. 
The Rips complex

Topologies of \( \{ f^{-1}([ -\infty, \alpha]) \} \) and \( \{ \tilde{f}^{-1}([ -\infty, \alpha]) \} \) are completely different:
The Rips complex

Topologies of \( \{ f^{-1}([-\infty, \alpha]) \} \) and \( \{ \tilde{f}^{-1}([-\infty, \alpha]) \} \) are completely different:
The Rips complex

Topologies of \( \{ f^{-1}([-\infty, \alpha]) \} \) and \( \{ \tilde{f}^{-1}([-\infty, \alpha]) \} \) are completely different:

\[
\begin{array}{cccccc}
  & \bullet & \bullet & \bullet & \bullet & \bullet \\
\end{array}
\]

Giving thickness by building a Rips complex:
The Rips complex

Topologies of \( \{ f^{-1}([-\infty, \alpha]) \} \) and \( \{ \tilde{f}^{-1}([-\infty, \alpha]) \} \) are completely different:

\[ \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \]

Giving thickness by building a Rips complex:
The Rips complex

Topologies of $\{f^{-1}(-\infty, \alpha)\}$ and $\{\tilde{f}^{-1}(-\infty, \alpha]\}$ are completely different:

\[ \bullet \bullet \bullet \bullet \bullet \]

Giving thickness by building a Rips complex:
The Rips complex

Topologies of \( \{ f^{-1}([-\infty, \alpha]) \} \) and \( \{ \tilde{f}^{-1}([-\infty, \alpha]) \} \) are completely different:

\[
\begin{array}{cccccc}
\bullet & \bullet & \bullet & \bullet & \bullet & \bullet
\end{array}
\]

Giving thickness by building a Rips complex:
The Rips complex

Topologies of \( \{ f^{-1}([\infty, \alpha]) \} \) and \( \{ \tilde{f}^{-1}(\infty, \alpha) \} \) are completely different:

Giving thickness by building a Rips complex:
Nested Rips complexes

Sometimes, there is no parameter such that the Rips complex capture the topology of $M$. 

Nested Rips complexes

Sometimes, there is no parameter such that the Rips complex capture the topology of M.

Solution: take two different parameters $\delta < \delta'$ and look at the image of $H_\ast(R_\delta)$ by the morphism induced by the inclusion $R_\delta \hookrightarrow R_{\delta'}$. 
Nested Rips complexes

Sometimes, there is no parameter such that the Rips complex capture the topology of $M$.

Solution: take two different parameters $\delta < \delta'$ and look at the image of $H_\ast(R_\delta)$ by the morphism induced by the inclusion $R_\delta \hookrightarrow R_{\delta'}$.

Effect of ”cleaning” the persistence diagram.
Filtration by functional values

We study the scalar field $f$ and not the manifold $M$. 

Filtration by functional values

We study the scalar field $f$ and not the manifold $M$.

We work for fixed parameters $\delta$ and $\delta'$ and we use the values of $\tilde{f}$ to build the filtration:

$$P_\alpha = \tilde{f}^{-1}([\minus\infty, \alpha])$$
Filtration by functional values

We study the scalar field $f$ and not the manifold $M$.

We work for fixed parameters $\delta$ and $\delta'$ and we use the values of $\tilde{f}$ to build the filtration:

$$P_\alpha = \tilde{f}^{-1}([-\infty, \alpha])$$

$$\{R_\delta(P_\alpha) \leftrightarrow R_{\delta'}(P_\alpha)\}_{\alpha \in \mathbb{R}}$$
Theoretical guarantees

From Chazal, Guibas, Oudot and Skraba (DCG’11)
Let $\rho(M)$ be the strong convexity radius of $M$. 

Theorem
If $P$ is an $\epsilon$-Riemannian sample of $M$, then:

$$||f_P - \tilde{f}||_\infty \leq \xi$$

and

$$\epsilon < \frac{1}{4} \rho(M): \forall \delta \in \left[2 \epsilon, \frac{1}{2} \rho(M)\right],$$

$$d_B(Dgm(f), Dgm(R_\delta(P_\alpha) \hookrightarrow R_\delta'(P_\alpha))) \leq 2c\delta + \xi$$

Theorem
If $P$ is an $\epsilon$-Riemannian sample of $M$ and the distance between points of $P$ are given by a function $\tilde{d}$ such that:

$$d_M(x, y) \leq \tilde{d}(x, y) \leq \nu + \mu d_M(x, y) \lambda,$$

then:

$$\forall \delta \geq \nu + 2\mu \epsilon \lambda, \delta' \in \left[\nu + 2\mu \delta, \frac{1}{\lambda} \rho(M)\right],$$

$$d_B(Dgm(f), Dgm(R_\delta(P_\alpha) \hookrightarrow R_\delta'(P_\alpha))) \leq c\lambda \delta'$$
Theoretical guarantees

From Chazal, Guibas, Oudot and Skraba (DCG’11)
Let $\varrho(M)$ be the strong convexity radius of $M$.

**Theorem**

If $P$ is an $\epsilon$ Riemannian sample of $M$, $\|f|_P - \tilde{f}\|_{\infty} \leq \xi$ and $\epsilon < \frac{1}{4}\varrho(M)$:

$$\forall \delta \in \left[2\epsilon, \frac{1}{2}\varrho(M)\right[, d_B(Dgm(f), Dgm(R_{\delta}(P_\alpha) \hookrightarrow R_{2\delta}(P_\alpha))) \leq 2c\delta + \xi$$
Theoretical guarantees

From Chazal, Guibas, Oudot and Skraba (DCG’11)
Let $\varrho(M)$ be the strong convexity radius of $M$.

**Theorem**

If $P$ is an $\epsilon$ Riemannian sample of $M$, $\|f|_P - \tilde{f}\|_\infty \leq \xi$ and $\epsilon < \frac{1}{4} \varrho(M)$:

$$\forall \delta \in [2\epsilon, \frac{1}{2} \varrho(M)], d_B(Dgm(f), Dgm(R_\delta(P_\alpha) \hookrightarrow R_{2\delta}(P_\alpha))) \leq 2c\delta + \xi$$

**Theorem**

If $P$ is an $\epsilon$ Riemannian sample of $M$ and the distance between points of $P$ are given by a function $\tilde{d}$ such that $\frac{d_M(x,y)}{\lambda} \leq \tilde{d}(x,y) \leq \nu + \mu \frac{d_M(x,y)}{\lambda}$, then:

$$\forall \delta \geq \nu + 2\mu \frac{\epsilon}{\lambda}, \delta' \in [\nu + 2\mu \delta, \frac{1}{\lambda} \varrho(M)],$$

$$d_B(Dgm(f), Dgm(R_\delta(P_\alpha) \hookrightarrow R_{\delta'}(P_\alpha))) \leq c\lambda\delta'$$
Sources of noise
Sources of noise
Sources of noise
Sources of noise
Sources of noise
Discrepancy

We compute new functionnal values for points:
Discrepancy

We compute new functionnal values for points:

For every point $p$ in $P$:
Discrepancy

We compute new functionnal values for points:

For every point $p$ in $P$:

1. Build the set $NN_k(p)$ of its $k$ nearest neighbors.
We compute new functionnal values for points:

For every point $p$ in $P$:

1. Build the set $NN_k(p)$ of its $k$ nearest neighbors.
2. Find the subset $Y$ of $k'$ values in $\tilde{f}(NN_k(p))$ with the smallest variance.
Discrepancy

We compute new functionnal values for points:

For every point $p$ in $P$:

1. Build the set $NN_k(p)$ of its $k$ nearest neighbors.

2. Find the subset $Y$ of $k'$ values in $\tilde{f}(NN_k(p))$ with the smallest variance.

3. Fix the new function value $\hat{f}(p)$ as the barycenter of $Y$. 
A variant using the median

We compute new functionnal values for points :

For every point $p$ in $P$:

1. Build the set $NN_k(p)$ of its $k$ nearest neighbors.

2. Fix the new function value $\hat{f}(p)$ as the median of $\tilde{f}(NN_k(p))$. 
Asymptotic behaviour

When $k \to \infty$ and $\frac{k}{n} \to 0$. 
Asymptotic behaviour

When $k \rightarrow \infty$ and $\frac{k}{n} \rightarrow 0$.

Probability distribution around the correct value:
Asymptotic behaviour

When $k \to \infty$ and $\frac{k}{n} \to 0$.

Probability distribution around the correct value:
Asymptotic behaviour

When $k \to \infty$ and $\frac{k}{n} \to 0$.

Probability distribution around the correct value:
Asymptotic behaviour

When $k \to \infty$ and $\frac{k}{n} \to 0$.

Probability distribution around the correct value:
Asymptotic behaviour

When $k \to \infty$ and $\frac{k}{n} \to 0$.

Probability distribution around the correct value:
Asymptotic behaviour

When \( k \to \infty \) and \( \frac{k}{n} \to 0 \).

Probability distribution around the correct value:
Asymptotic behaviour

When $k \to \infty$ and $\frac{k}{n} \to 0$.

Probability distribution around the correct value:
Asymptotic behaviour

When \( k \to \infty \) and \( \frac{k}{n} \to 0 \).

Probability distribution around the correct value:
Asymptotic behaviour

When $k \to \infty$ and $\frac{k}{n} \to 0$.

Probability distribution around the correct value:
**Bone**

<table>
<thead>
<tr>
<th></th>
<th>Noisy input</th>
<th>k-NN regression</th>
<th>Discrepancy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max</td>
<td>16.23</td>
<td>3.18</td>
<td>0.37</td>
</tr>
<tr>
<td>Mean</td>
<td>0.349</td>
<td>0.204</td>
<td>0.097</td>
</tr>
</tbody>
</table>
Experimental illustration

Persistence of topographic map

Topographic map

Original persistence diagram

SNR = 2.96
Persistence of topographic map

Topographic map  Noisy topographic map

Original persistence diagram  SNR=2.96
Persistençe of topographic map

Topographic map

Noisy topographic map

Noisy persistence diagram

Original persistence diagram
Persistence of topographic map

Topographic map

Noisy topographic map

Noisy persistence diagram

Original persistence diagram

$k$-NN persistence diagram

Discrepancy persistence diagram
Images

No noise
Images

No noise

40% outliers
Images

No noise

40% outliers

kNN
Experimental illustration

Images

No noise

40% outliers

kNN

Discrepancy

Median
Experimental illustration

Lena’s diagrams
Lena’s diagrams
Lena’s diagrams
Lena’s diagrams
Chessboard

No noise

30% outliers
**Chessboard**

![Chessboard images](image_url)

- No noise
- 30% outliers
- kNN
- Discrepancy
- Median
Building a complex with good properties

We need a complex that has the correct geometric structure to analyze the scalar field.
Building a complex with good properties

We need a complex that has the correct geometric structure to analyze the scalar field.

- Use of the super-level sets of a density estimator
- Use of the sub-level sets of a distance-like function.
Distance to measure in a nutshell

The distance to a measure is a distance-like function designed to cope with outliers.

\[ d_{\mu,m}(p) = \sqrt{\frac{1}{k} \sum_{x \in \mathcal{NN}_k(p)} ||p - x||^2} \]
Distance to measure in a nutshell

The distance to a measure is a distance-like function designed to cope with outliers.

\[ d_{\mu,m}(p) = \sqrt{\frac{1}{k} \sum_{x \in NN_k(p)} ||p - x||^2} \]

- Easy to compute pointwise.
Geometric denoising

Distance to measure in a nutshell

The distance to a measure is a distance-like function designed to cope with outliers.

\[ d_{\mu,m}(p) = \sqrt{\frac{1}{k} \sum_{x \in NN_k(p)} \|p - x\|^2} \]

- Easy to compute pointwise.
- Guarantees of geometric inference
  [Chazal, Cohen-Steiner, Mérigot, 2011]
Distance to measure in a nutshell

The distance to a measure is a distance-like function designed to cope with outliers.

\[ d_{\mu,m}(p) = \sqrt{\frac{1}{k} \sum_{x \in NN_k(p)} \| p - x \|^2} \]

- Easy to compute pointwise.
- Guarantees of geometric inference
  [Chazal, Cohen-Steiner, Mérigot, 2011]
- Properties of a density estimator
  [Biau, Chazal, Cohen-Steiner, Devroye, Rodrigues, 2011]
A complete noise model

Three conditions:
A complete noise model

Three conditions:

1. Dense sampling:

\[ \forall x \in M, \ d_{\mu,m}(x) \leq \epsilon \]
A complete noise model

Three conditions:

1. Dense sampling:

$$\forall x \in M, \ d_{\mu,m}(x) \leq \epsilon$$

2. No cluster of noise:

$$r = \sup \{ l \in \mathbb{R} | \forall x, \ d_{\mu,m}(x) < l \implies d(x, M) \leq d_{\mu,m}(x) + \epsilon \}$$
A complete noise model

Three conditions:

1. Dense sampling:

   \[ \forall x \in M, \quad d_{\mu,m}(x) \leq \epsilon \]

2. No cluster of noise:

   \[ r = \sup \{ l \in \mathbb{R} | \forall x, \quad d_{\mu,m}(x) < l \implies d(x,M) \leq d_{\mu,m}(x) + \epsilon \} \]

3. For any point close to \( M \), most of the neighboring values are good:

   \[ \forall p \in d_{\mu,m}^{-1}([-\infty, \eta]), \quad |\{ q \in NN_k(p) | |\tilde{f}(q) - f(\pi(p))| \leq s \}| \geq k' \]
Theoretical results

Theorem

If $P$ is a set verifying the previous conditions and $f$ is a $c$-Lipschitz function then:

$$\forall \delta \in [2\eta + 6\epsilon, \frac{\phi(M)}{2}], \quad \delta' \in [2\eta + 2\epsilon + \frac{2R_M}{R_M - (\eta + \epsilon)} \delta, \frac{R_M - (\eta + \epsilon)}{R_M} \phi(M)],$$

$$d_B(Dgm(f), \hat{D}) \leq \left( \frac{cR_M \delta'}{R_M - (\eta + \epsilon)} + \xi s \right)$$

with $\xi = 1$ for the median and $\xi = 1 + 2\sqrt{\frac{k-k'}{2k'-k}}$ for the discrepancy.
Take home

A versatile and model free algorithm for functional denoising
 Scalar field analysis with noise in both the geometry and the functional values

But...

The algorithm needs some parameters.

Heuristics exist but there is no general method to choose their value.
Take home

- A versatile and model free algorithm for functional denoising
Take home

- A versatile and model free algorithm for functional denoising
- Scalar field analysis with noise in both the geometry and the functional values
Take home

- A versatile and model free algorithm for functional denoising
- Scalar field analysis with noise in both the geometry and the functional values

But...
Take home

- A versatile and model free algorithm for functional denoising
- Scalar field analysis with noise in both the geometry and the functional values

But...

- The algorithm needs some parameters.
Take home

- A versatile and model free algorithm for functional denoising
- Scalar field analysis with noise in both the geometry and the functional values

But...

- The algorithm needs some parameters.
- Heuristics exist but there is no general method to choose their value.