Machine Learning with Topological Data Analysis features

Martina Scolamiero

Online Machine Learning Seminar,
8 November 2023

Plan

- Stable ranks: stable topological feature maps for (noisy) data.

  J. Agerberg, W. Chachólski, R. Ramanujam, based on work of O. Gävfert and H. Riihimaki.

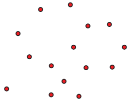
- Microglia MorphOMICs detects microglia morphological signatures across conditions.

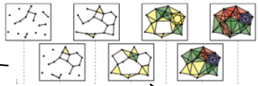
  G. Colombo, R. Cubero, L. Kanari, A. Venturino, R. Shulz, M. Scolamiero, J. Agerberg, H. Mathys, L.Tsai, W. Chachólski, K. Hess, S. Siegert.

- Paramterized families of stable ranks.

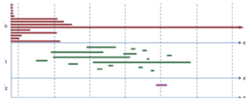
  A. Guidolin, J. Agerberg, I. Ren.

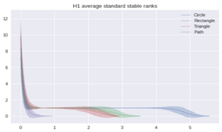
Persistence Pipeline

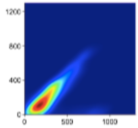




$$V : V_0 \to V_1 \to \ldots \to V_n$$







Persistence Modules

Let K be a field. A persistence module is a functor

$$\begin{aligned} F : \mathbb{R} &\longrightarrow \mathrm{Vect}_K \\ v &\mapsto F(v) \\ v \leq w &\mapsto F(v) \to F(w). \end{aligned}$$

Persistence Modules

Let K be a field. A persistence module is a functor

$$\begin{aligned} F : \mathbb{R} &\longrightarrow \mathrm{Vect}_K \\ v &\mapsto F(v) \\ v \le w &\mapsto F(v) \to F(w). \end{aligned}$$

A morphism of persistence modules $\phi : F \to G$ is given by a linear maps $\phi(v) : F(v) \to G(v)$ s.t if $v \le w$ the following diagram commutes

$$\begin{tikzcd} \mathrm{F(v)} \arrow[r, "F(v \le w)"] \arrow[d, "\phi(v)"'] & F(W) \arrow[d, "\phi(w)"] \\ G(v) \arrow[r, "G(v \le w)"] & G(w) \end{tikzcd}$$

$\mathcal{T}$ is the category of tame (i.e finitely generated) persistence modules and morphisms between them.

Multi-Persistence Modules

More generally we can consider persistence modules as functors of the form

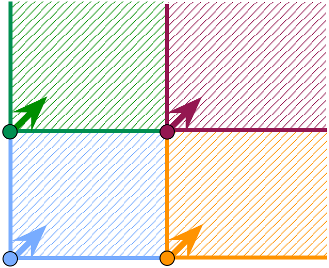
$$\begin{aligned} F : \mathbb{R}^r &\longrightarrow \mathrm{Vect}_K \\ v &\mapsto F(v) \\ v \leq w &\mapsto F(v) \to F(w). \end{aligned}$$

Multi-Persistence Modules

More generally we can consider persistence modules as functors of the form

$$\begin{aligned} F : \mathbb{R}^r &\longrightarrow \mathrm{Vect}_K \\ v &\mapsto F(v) \\ v \leq w &\mapsto F(v) \to F(w). \end{aligned}$$

and assume they are tame



Rank

- Tame persistence modules have finite projective dimension;
- Multigraded Betti numbers can be computed $\{\beta_{i,v}\}_{v \in I_i}$.

We focus on the rank of a tame functor $F \colon \mathbb{R}^r \to \mathrm{Vect}_K$

$$\operatorname{rank} F := \sum_{v \in I_0} \beta_{0,v}$$

Stable Rank

Let d be a distance on $\mathcal{T}$ and $\text{rank} : \mathcal{T} \to \mathbb{N}$ associating to a persistence module its rank.

The stable rank along d is the function:
$\widehat{\text{rank}}_d : \mathcal{T} \to \text{Fun}([0,\infty),[0,\infty))$ with values:

$$\widehat{\text{rank}}_d(F)(t) := \begin{cases} \text{rank}(F) & \text{if } t = 0 \\ \inf\{\text{rank}(G) \mid d(F,G) \leq t\} & \text{if } t > 0 \end{cases}$$

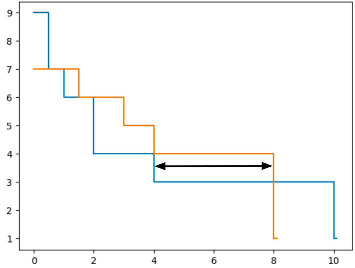
The idea of hierarchical stabilization is similar to the one of persistence signatures (Bauer et al.)

Persistence Barcodes Versus Kolmogorov Signatures: Detecting Modes of One-Dimensional Signals, U.Bauer, A. Munk, H. Sieling, M. Wardetzky, Foundations of Computational Mathematics (2017).

Stability

Two functions $f, g: [0, \infty) \to [0, \infty)$ are ϵ-interleaved if:

$$f(v) \geq g(v+\epsilon) \quad \text{and} \quad g(v) \geq f(v+\epsilon), \quad \forall v \in [0, \infty)$$



The interleaving distance between two functions is:

$$d_I(f,g) := \inf\{\epsilon \mid f, g \text{ are } \epsilon - \text{interleaved}\}$$

Stability

The function $\widehat{\text{rank}}_d : (\mathcal{T}, d) \to (\text{Fun}([0,\infty),[0,\infty)), d_I)$ is 1-Lipschitz:

$$\forall F, G \in (\mathcal{T}, d), \quad d_I(\widehat{\text{rank}}_d(F), \widehat{\text{rank}}_d(G)) \leq d(F, G)$$

Noise systems

For $\epsilon \in [0, \infty)$, noise of size ϵ is a set $\mathcal{S}_\epsilon$ of persistence modules

The collection $\{\mathcal{S}_\epsilon\}_{\epsilon \in [0,\infty)}$ should satisfy the conditions:

- the module 0 belongs to $\mathcal{S}_\epsilon$ for any ϵ;

Multidimensional Persistence and Noise M. Scolamiero,W. Chacholski, R. Ramanujam, A. Lundman, S. Oberg. Found Comput Math (2017).

Noise systems

For $\epsilon \in [0, \infty)$, noise of size ϵ is a set $\mathcal{S}_\epsilon$ of persistence modules

The collection $\{\mathcal{S}_\epsilon\}_{\epsilon \in [0,\infty)}$ should satisfy the conditions:

- the module 0 belongs to $\mathcal{S}_\epsilon$ for any ϵ;
- if $0 \leq \tau \leq \epsilon$, then $\mathcal{S}_\tau \subseteq \mathcal{S}_\epsilon$;

Multidimensional Persistence and Noise M. Scolamiero,W. Chacholski, R. Ramanujam, A. Lundman, S. Oberg. Found Comput Math (2017).

Noise systems

For $\epsilon \in [0, \infty)$, noise of size ϵ is a set $\mathcal{S}_\epsilon$ of persistence modules

The collection $\{\mathcal{S}_\epsilon\}_{\epsilon \in [0,\infty)}$ should satisfy the conditions:

- the module 0 belongs to $\mathcal{S}_\epsilon$ for any ϵ;
- if $0 \leq \tau \leq \epsilon$, then $\mathcal{S}_\tau \subseteq \mathcal{S}_\epsilon$;
- if $0 \to F \to G \to H \to 0$ is an exact sequence, then
  - if G is in $\mathcal{S}_\epsilon$, then so are F and H;
  - if F is in $\mathcal{S}_\epsilon$ and H is in $\mathcal{S}_\tau$, then G is in $\mathcal{S}_{\epsilon+\tau}$.

Multidimensional Persistence and Noise M. Scolamiero, W. Chacholski, R. Ramanujam, A. Lundman, S. Oberg. Found Comput Math (2017).

Standard noise in the direction of a cone $\mathcal{V} \subseteq \mathbb{R}^r$

$$\mathcal{V}_\epsilon := \left\{ F \text{ tame} \mid \begin{array}{c} \text{for any } x \text{ in } F(v)\text{, there is } w \text{ in } \mathrm{Cone}(\mathcal{V}) \text{ s.t.} \\ \|w\|_\infty = \epsilon \text{ and } x \text{ is in } \ker\left(F(v) \to F(v+w)\right) \end{array} \right\}$$

Standard noise in the direction of a cone $\mathcal{V} \subseteq \mathbb{R}^r$

$$\mathcal{V}_\epsilon := \left\{ F \text{ tame} \mid \begin{array}{c} \text{for any } x \text{ in } F(v) \text{, there is } w \text{ in Cone}(\mathcal{V}) \text{ s.t.} \\ \|w\|_\infty = \epsilon \text{ and } x \text{ is in } \ker\left(F(v) \to F(v+w)\right) \end{array} \right\}$$

Given the cone $\mathcal{V}$ generated by $(1,1) \in \mathbb{R}^2$, the following 1-tame functor is in $\mathcal{V}_1$

$$\begin{tikzcd}
k \arrow[r] & 0 \arrow[r] & 0 \\
k \arrow[r] \arrow[u] & 0 \arrow[r] \arrow[u] & 0 \arrow[u] \\
k \arrow[r] \arrow[u] & k \arrow[r] \arrow[u] & k \arrow[u]
\end{tikzcd}$$

- Volume noise system

Given a tame functor F, consider $\mathrm{dom}(F) := \{v \in \mathbb{R}^r \mid F(v) \neq 0\}$

$$V_\epsilon := \{F \in \mathcal{T} \mid \mathrm{vol}(\mathrm{dom}(F)) \leq \epsilon\}.$$

Master thesis of Daniel Lundin.

- Volume noise system

Given a tame functor F, consider $\mathrm{dom}(F) := \{v \in \mathbb{R}^r \mid F(v) \neq 0\}$

$$V_\epsilon := \{F \in \mathcal{T} \mid \mathrm{vol}(\mathrm{dom}(F)) \leq \epsilon\}.$$

- Hilbert noise system

$$\mathcal{H}_\epsilon := \{F \in \mathcal{T} \mid \int_{\mathrm{dom(F)}} \dim_K F(v) \leq \epsilon\}.$$

• r=1 p-norms

For $p \in [1, \infty]$ consider the p-norm of a persistence module $F \simeq \bigoplus_{i=1}^{k} [b_i, d_i)$ defined as:

$$\|F\|_p := \begin{cases} \left(\sum_{i=1}^{k} |d_i - b_i|^p\right)^{\frac{1}{p}} & \text{if } p \neq \infty \\ \max_{i=1}^{k} |d_i - b_i| & \text{if } p = \infty \end{cases}$$

- The sets $\mathcal{S}_\epsilon^p := \{F \in \mathcal{T} \mid \quad \|F\|_p \leq \epsilon\}$ form a noise system.

Wasserstein Stability for Persistence Diagrams P. Skraba, K. Turner arXiv:2006.16824v3.

Algebraic Wasserstein distances and stable homological invariants of data, J. Agergerg, A. Guidolin, I. Ren, M. Scolamiero arXiv:2301.06484.

• r=1 contours

A contour is a function $C: [0,\infty] \times [0,\infty) \to [0,\infty]$ s.t:

- $C(a,0) = a$
- $C(C(a,\epsilon),\tau) = C(a,\epsilon+\tau)$
- $C(a,\epsilon) \leq C(b,\tau)$ if $a \leq b, \quad \epsilon \leq \tau$

Metrics and Stabilization in One Parameter Persistence, W.Chacholski, H. Riihimaki SIAM J. ApplL. Algebra Geometry 2020

- r=1 contours

A contour is a function $C : [0,\infty] \times [0,\infty) \to [0,\infty]$ s.t:

- $C(a,0) = a$
- $C(C(a,\epsilon),\tau) = C(a,\epsilon+\tau)$
- $C(a,\epsilon) \leq C(b,\tau)$ if $a \leq b, \quad \epsilon \leq \tau$

Each contour defines a noise system with components:

$$\mathcal{C}_\epsilon := \left\{ F \text{ tame} \mid \text{ for any } v, F(v) \to F(C(v,\epsilon)) \text{ is the zero map } \right\}$$

Metrics and Stabilization in One Parameter Persistence, W.Chacholski, H. Riihimaki SIAM J. ApplL. Algebra Geometry 2020

Contours

A regular contour is a function $C\colon [0,\infty]\times[0,\infty)\to[0,\infty]$ s.t:

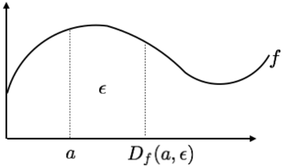
- $C(a,0)=a$
- $C(C(a,\epsilon),\tau)=C(a,\epsilon+\tau)$
- $C(a,\epsilon)\leq C(b,\tau)$ if $a\leq b, \quad \epsilon\leq\tau$
- $C(-,\epsilon):[0,\infty]\to[0,\infty]$ is a monomorphism $\forall\epsilon\in[0,\infty)$
- $C(a,-):[0,\infty)\to[0,\infty]$ is a monomorphism whose image is $[a,\infty]$, for every $a\in[0,\infty)$.

Metrics and Stabilization in One Parameter Persistence, W.Chacholski, H. Riihimaki SIAM J. ApplL. Algebra Geometry 2020

Densities

Each measurable function $f : [0,\infty) \to (0,\infty)$ defines a contour:

$$\begin{aligned} D_f : [0,\infty)\times[0,\infty) &\longrightarrow [0,\infty) \\ (a,\epsilon) &\mapsto D_f(a,\epsilon) \quad s.t \int_a^{D_f(a,\epsilon)} f(x)\,dx = \epsilon \end{aligned}$$



If f is constant with value 1, then $D_f(a,\epsilon) = a + \epsilon$.

Lifespan

For a regular contour $C\colon [0,\infty]\times[0,\infty)\to[0,\infty]$ the lifespan of a bar $[b,d)$ is the unique non-negative real number ℓ such that $C(b,\ell)=d$

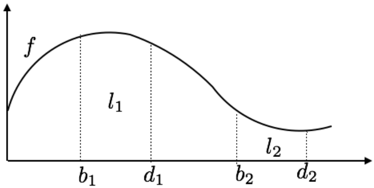
If $C(a,\epsilon):=a+\epsilon$, the lifespan of $[b,d)$ is its length $d-b$.

Lifespan

For a regular contour $C\colon [0,\infty] \times [0,\infty) \to [0,\infty]$ the lifespan of a bar $[b,d)$ is the unique non-negative real number ℓ such that $C(b,\ell) = d$

If $C(a,\epsilon) := a + \epsilon$, the lifespan of $[b,d)$ is its length $d - b$.

With density type contours D_f, we can differently weigh parts of the parameter space.



p,C-norms

For $p \in [1, \infty]$ and C a regular contour, the (p, C)-norm of a persistence module $F \simeq \bigoplus_{i=1}^{k} [b_i, d_i)$ is defined as:

$$\|F\|_{p,C} := \begin{cases} \left(\sum_{i=1}^{k} \ell_i^p\right)^{\frac{1}{p}} & \text{if } p \neq \infty \\ \max\{\ell_i\}_{i=1}^{k} & \text{if } p = \infty \end{cases}$$

where ℓ_i denotes the lifespan of $[b_i, d_i)$ with respect to C.

p,C-norms

For $p \in [1, \infty]$ and C a regular contour, the (p, C)-norm of a persistence module $F \simeq \bigoplus_{i=1}^{k} [b_i, d_i)$ is defined as:

$$\|F\|_{p,C} := \begin{cases} \left(\sum_{i=1}^{k} \ell_i^p\right)^{\frac{1}{p}} & \text{if } p \neq \infty \\ \max\{\ell_i\}_{i=1}^{k} & \text{if } p = \infty \end{cases}$$

where ℓ_i denotes the lifespan of $[b_i, d_i)$ with respect to C.

- The sets $\mathcal{S}_{\epsilon}^{p,C} := \{F \in \mathcal{T} \mid \quad \|F\|_{p,C} \leq \epsilon\}$ form a noise system.

A noise system $\{\mathcal{S}_\epsilon\}_{\epsilon \in \mathbb{Q}}$ is closed under direct sums if:

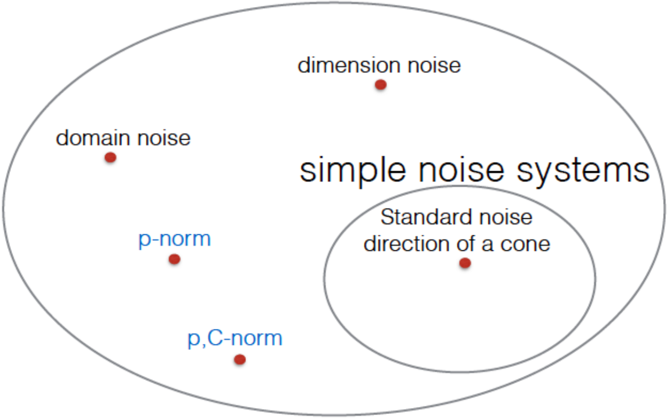
$$\text{for every } X, Y \in \mathcal{S}_\epsilon \Rightarrow X \oplus Y \in \mathcal{S}_\epsilon$$

A noise system $\{\mathcal{S}_\epsilon\}_{\epsilon \in \mathbb{Q}}$ is closed under direct sums if:

$$\text{for every } X, Y \in \mathcal{S}_\epsilon \Rightarrow X \oplus Y \in \mathcal{S}_\epsilon$$

- Contour noise systems are closed under direct sums while p-norm and volume noise systems are not.

Examples of noise systems



Pesudo-Metrics

A Noise System $\mathcal{S} := \{\mathcal{S}_\epsilon\}_{\epsilon\in[0,\infty)}$ induces a pseudo-metric on $\mathcal{T}$.

F,G : $\mathbb{R} \to \mathrm{Vect}_K$ are ϵ-close w.r.t $\mathcal{S}$ if and only if :

$$\begin{tikzcd} & H \arrow[dl, "\phi"'] \arrow[dr, "\psi"] & \\ F & & G \end{tikzcd}$$

$\ker\phi \in \mathcal{S}_{\epsilon_1}$, $\mathrm{coker}\phi \in \mathcal{S}_{\epsilon_2}$, $\ker\psi \in \mathcal{S}_{\epsilon_3}$, $\mathrm{coker}\,\psi \in \mathcal{S}_{\epsilon_4}$ and $\epsilon_1 + \epsilon_2 + \epsilon_3 + \epsilon_4 \leq \epsilon$.

$$d_{\mathcal{S}}(F,G) := \inf\{\epsilon\,|F,G\,\text{are}\,\epsilon-\text{close}\}$$

p-Wasserstein distance

Consider $\| \quad \|_p$, the p-norm in $\mathbb{R}^4$.

F,G : $\mathbb{R} \to \mathrm{Vect}_K$ are ϵ-close w.r.t $\mathcal{S}$ if and only if :

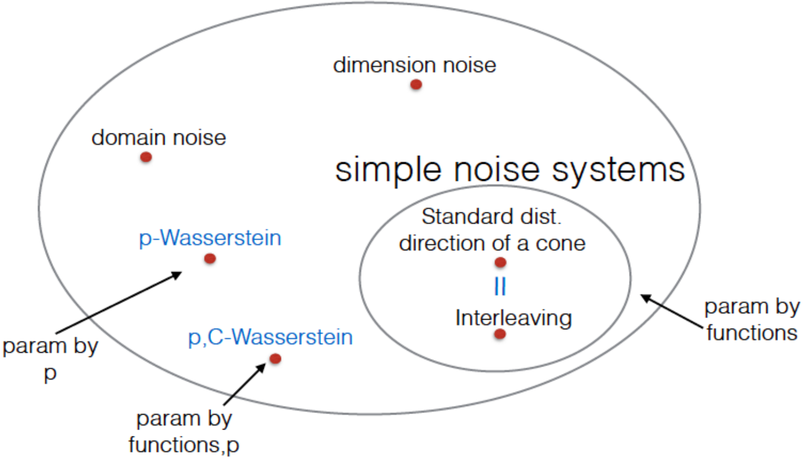
$$\begin{tikzcd} & H \arrow[dl, "\phi"'] \arrow[dr, "\psi"] & \\ F & & G \end{tikzcd}$$

$\ker \phi \in \mathcal{S}_{\epsilon_1}$, coker$\phi \in \mathcal{S}_{\epsilon_2}$, $\ker \psi \in \mathcal{S}_{\epsilon_3}$, coker $\psi \in \mathcal{S}_{\epsilon_4}$ and $\|\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4\|_p \leq \epsilon$.

$$d^p_{\mathcal{S}}(F,G) := \inf\{\epsilon \,|F, G \,\text{are}\, \epsilon - \text{close}\}$$

$d^p_{\mathcal{S}^p}$ coincides with the algebraic p-Wasserstein distance

Examples of Pseudometrics



Stable Rank Kernel

For X and Y in $\mathcal{T}$ the stable rank kernel w.r.t d is:

$$K_d(X,Y) := \int_0^\infty \widehat{\mathrm{rank}}_d(X)\widehat{\mathrm{rank}}_d(Y)dx$$

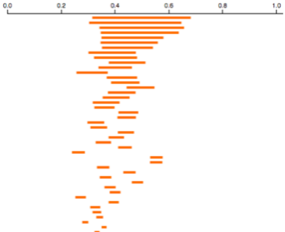
Stable Rank Kernel

For X and Y in $\mathcal{T}$ the stable rank kernel w.r.t d is:

$$K_d(X, Y) := \int_0^\infty \widehat{\text{rank}}_d(X)\widehat{\text{rank}}_d(Y)dx$$

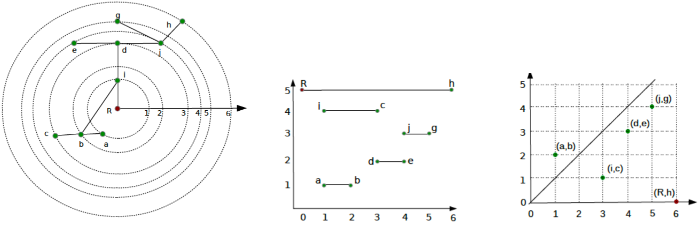
For contour noise systems, stable rank is a bar count: for $F \simeq \bigoplus_{i=1}^n [b_i, d_i)$:

$$\widehat{\text{rank}}_{\mathcal{S}}(F)(t) := |\{i \,|\, \ell_i > t\}|$$



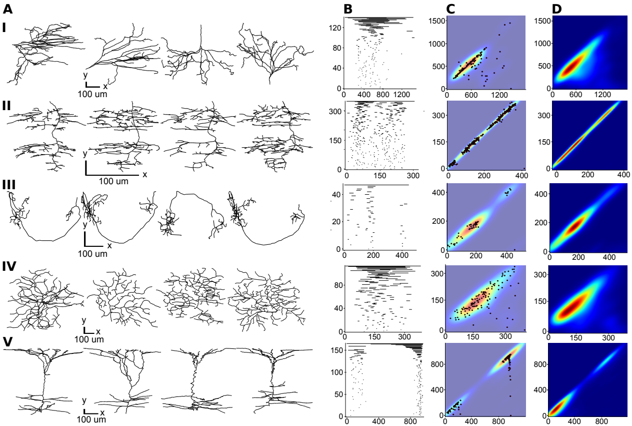
Topological Morphology Descriptor

Input: a rooted tree $T = (N, E)$ and a function $f : N \to \mathbb{R}$.
Output: barcode or persistence diagram.



Kanari et al. A topological representation of branching neuronal morphologies, Neuroinform, (2017)

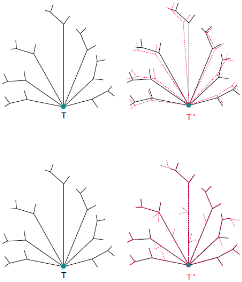
Neuronal Morphologies from different species



TMD representation of neurons from different animal species: cat, dragonfly, fruit fly, mouse and rat.

Stability of TMD

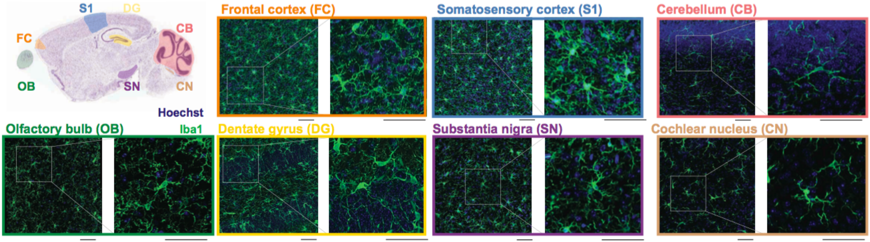
The TMD algorithm is stable with respect to: error in measuring the exact position of a node and omission or addition of a small branch.



One has to also account for intrinsic variability in the data!

Microglia Morphologies

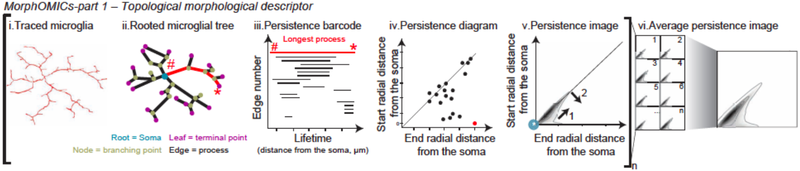
S. Siegert's lab at IST has been collecting and analysing almost 40,000 reconstructed microglia morphologies from mouse.



Microglial MorphOMICs unravel region- and sex-dependent morphological signatures from murine postnatal development to degeneration, G. Colombo, R. Cubero, L.Kanari . . . , W. Chacholski, K. Hess, S. Siegert.(Accepted: Nature Neuroscience)

Microglia MorphOmics

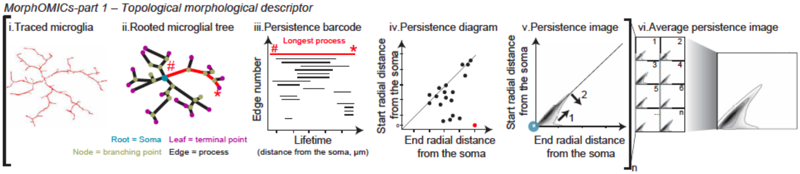
The first step in Morphomics is to compute the TMD of microglia trees.



Microglial MorphOMICs unravel region- and sex-dependent morphological signatures from murine postnatal development to degeneration, G. Colombo, R. Cubero, L.Kanari . . . , W. Chacholski, K. Hess, S. Siegert. Nature Neuroscience 2022.

Microglia MorphOmics

The first step in Morphomics is to compute the TMD of microglia trees.

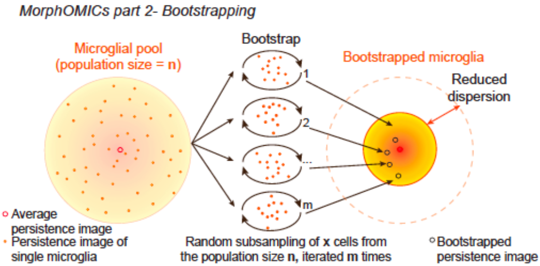


Because of intrinsic variability within the data, average persistence images or average stable ranks do not clearly exhibit regional signatures.

Microglial MorphOMICs unravel region- and sex-dependent morphological signatures from murine postnatal development to degeneration, G. Colombo, R. Cubero, L.Kanari . . . , W. Chacholski, K. Hess, S. Siegert. Nature Neuroscience 2022.

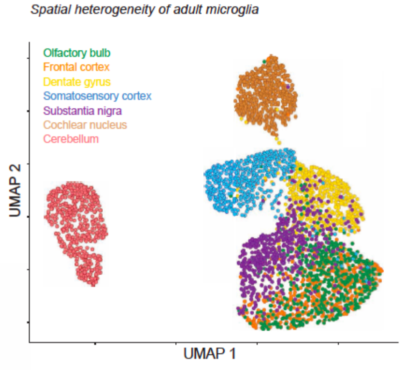
Bootstrap samples

The second step in Morphomics is to compute bootstrap samples of the data.



Dimensionality reduction

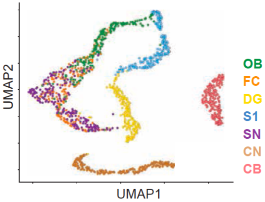
Bootstrap samples are then visualized through dimensionality reduction.



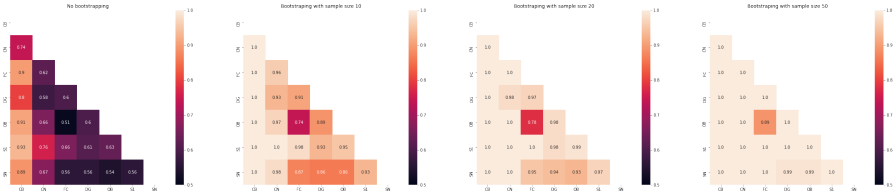
Boostrap samples of persistence images allow to identify spacial heterogeneity among adult microglia.

Classification with stable rank kernels

Standard stable ranks yield a similar visualization.



To be compared to accuracy in classification per brain region with stable rank kernel.



Computation of the p, C-Wasserstein stable rank

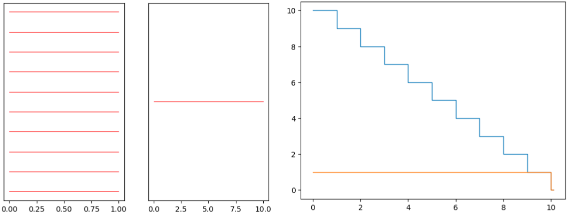
- $\widehat{\mathrm{rank}}_{p,C}(F)(t)$ is the minimal rank among all G that fit in a mono-epi span $F \hookleftarrow H \twoheadrightarrow G$ of bar to bar morphisms s.t $d(F, G) \leq t$.

Computation of the p, C-Wasserstein stable rank

- $\widehat{\text{rank}}_{p,C}(F)(t)$ is the minimal rank among all G that fit in a mono-epi span $F \hookleftarrow H \twoheadrightarrow G$ of bar to bar morphisms s.t $d(F, G) \leq t$.
- For each monomorphims $\phi : F \hookrightarrow H$ there exists a bar-to-bar monomorphism $\bar{\phi} : F \hookrightarrow H$ s.t $||\text{coker}\bar{\phi}||_{p,C} \leq ||\text{coker}\phi||_{p,C}$. Dually for epimorphisms morphisms can be assumed to be bar-to-bar.

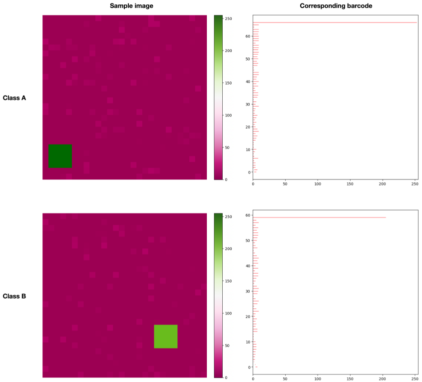
The p, C-Wasserstein stable rank

- Consider a barcode decomposition of $F = \bigoplus_{i=1}^{N} [b_i, d_i)$ s.t $\ell_1 \leq \ell_2 \leq \dots \ell_N$ where $C(b_i, \ell_i) = d_i$.
- The points $0 = t_0 < t_1 < \dots t_N$ of discontinuity of $\widehat{\mathrm{rank}}_p(F)$ are $t_i = 2^{\frac{1-p}{p}} \left(\sum_{j=1}^{i} \ell_j^p \right)^{\frac{1}{p}}$
- $\widehat{\mathrm{rank}}_p(F)(t_i) = N - i$



Algebraic Wasserstein distances and stable homological invariants of data, J. Agergerg, A. Guidolin, I. Ren, M. Scolamiero arXiv:2301.06484.

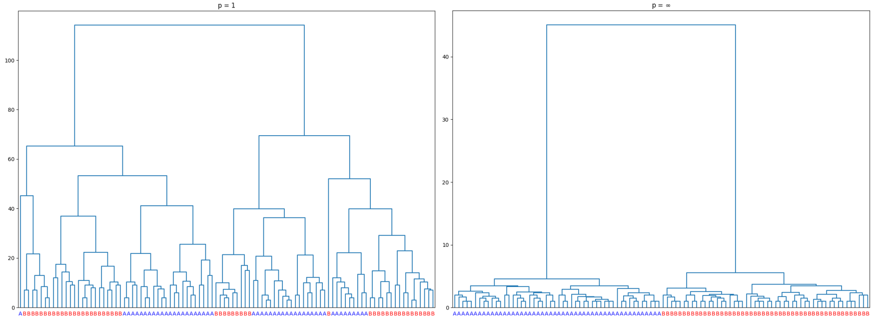
Signal vs Noise-artificial example I



High intensity pixels in images from class A are uniformly distributed between 245 and 255, in images from class B between 200 and 210. Low intensity pixels follow the same distribution in both classes.

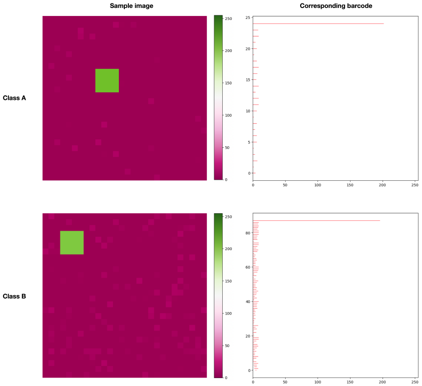
Signal vs Noise-artificial example I

Hierarchical clustering on p-Wasserstein stable ranks:
$p = 1(\mathit{left})$, $p = \infty$ (right).



$p = \infty$ is more appropriate in distinguishing different distributions of high intensity pixels.

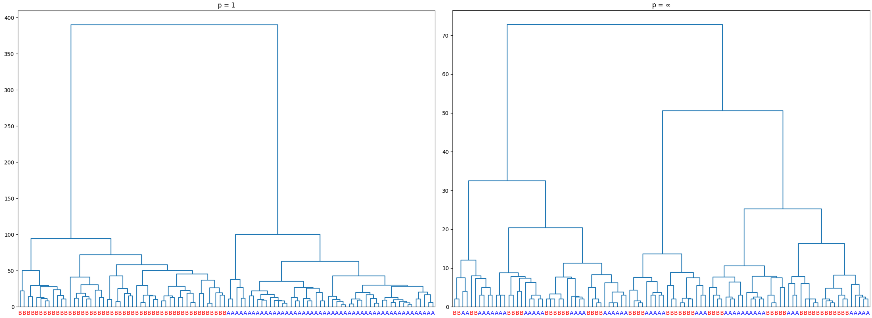
Signal vs Noise-artificial example II



The number of low intensity pixels in images from class A is uniformly distributed between 20 and 30, in images from class B between 120 and 130. High intensity pixels follow the same distribution in both classes.

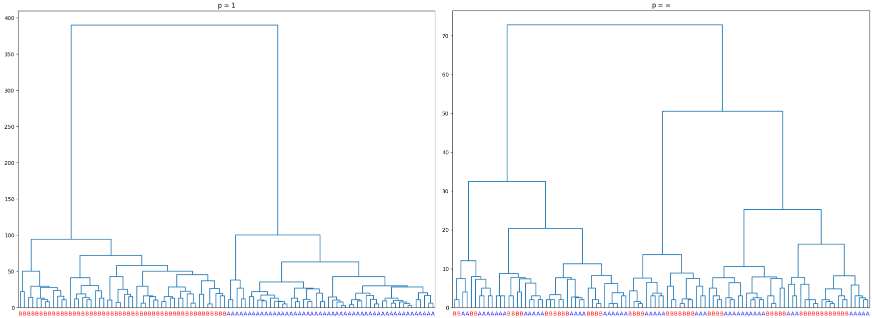
Signal vs Noise-artificial example II

Hierarchical clustering on p-Wasserstein stable ranks: $p = 1$(*left*), $p = \infty$ (right).



Signal vs Noise-artificial example II

Hierarchical clustering on p-Wasserstein stable ranks: $p = 1(\mathit{left})$, $p = \infty$ (right).



$p = 1$ is more appropriate in distinguishing different distributions of the number of low intensity pixels.

Dataset of brain artery trees

The dataset consists in 3D reconstructions of brain artery trees of 98 individuals of age $18 - 72$. The structure of brain arteries was found to significantly correlate with age.

Persistent homology analysis of brain artery trees. Paul Bendich, James S Marron, Ezra Miller, Alex Pieloch, and Sean Skwerer. The annals of applied statistics, 10(1):198, 2016.

Dataset of brain artery trees

The dataset consists in 3D reconstructions of brain artery trees of 98 individuals of age $18 - 72$. The structure of brain arteries was found to significantly correlate with age.

Classification task:

- We partitioned the indexing set I of the data, based on age of the subjects: class A (age < 45) and class B (age ≥ 45).
- Do stable ranks $\widehat{\text{rank}}_d(X_i)$ associated to a Wassertein-contour metric $d = d_{p,C}$ provide a good classification?
- How to learn p and the parameters associated to C?

Persistent homology analysis of brain artery trees. Paul Bendich, James S Marron, Ezra Miller, Alex Pieloch, and Sean Skwerer. The annals of applied statistics, 10(1):198, 2016.

Metric Learning

We used a metric learning objective function, designed to yield small intra-class distances and large inter-class distances:

$$\text{obj} = \frac{\sum_{i,j \in A} D^2(X_i, X_j)}{\sum_{i \in A, j \in I} D^2(X_i, X_j)} + \frac{\sum_{i,j \in B} D^2(X_i, X_j)}{\sum_{i \in B, j \in I} D^2(X_i, X_j)},$$

where $D(X_i, X_j)$ denotes the interleaving distance between the stable ranks of X_i and X_j.

The objective function is parametrized by p and by the parameters of the density

$$f(x) = \mathcal{N}(x|\mu_1, \sigma_1) + \lambda \mathcal{N}(x|\mu_2, \sigma_2)$$

that defines the contour.

Topology-based metric learning. Oliver Gävfert. Poster 2018.

Algebraic Wasserstein distances and stable homological invariants of data, J. Agergerg, A. Guidolin, I. Ren, M. Scolamiero arXiv:2301.06484.

Qualitative understanding of optimization

Results for one example run of the metric learning optimization over 25000 iterations.

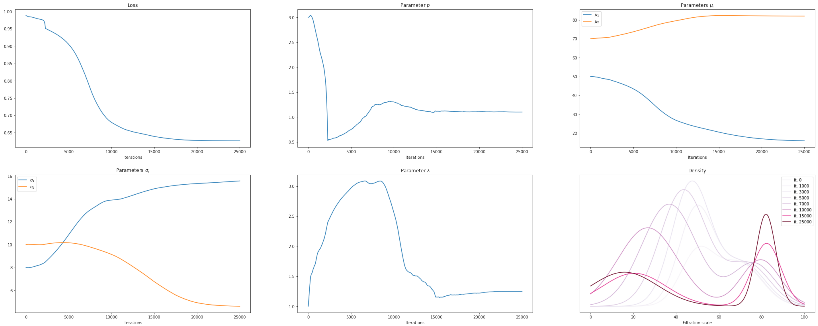


Figure: Progression of the loss, Wasserstein p parameter, mean μ_i, standard deviation σ_i and λ over the iterations. Density at different iterations.

Understanding the learnt density

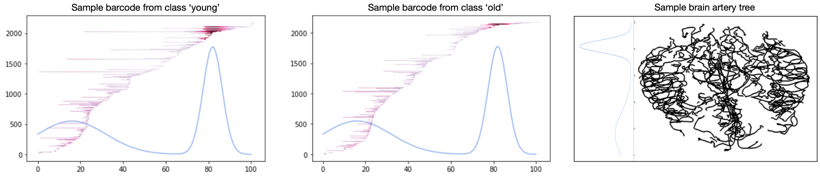


Figure: **Left and Middle:** Sample barcodes from the two classes with superposed density. Bars are colored according to the density. **Right:** Sample brain artery tree with superposed density.

The optimized stable rank is used as a classifier with a KNN method yielding a classification error of 24%. Before optimization, the classification error varied between 28.9% and 55%. Standard stable rank yields classification error of 38.1%.

In Summary

- Stable ranks are well suited for statistics and machine learning, are used in real world applications.

In Summary

- Stable ranks are well suited for statistics and machine learning, are used in real world applications.
- Varying the metric defining a stable rank, offers multiple perspectives on the data.

In Summary

- Stable ranks are well suited for statistics and machine learning, are used in real world applications.
- Varying the metric defining a stable rank, offers multiple perspectives on the data.
- Stable ranks provide stable features with respect to several metrics including p-Wasserstein distances on barcodes.

Thank you for your attention!