Persistent homology of stochastic processes and their zeta functions

PhD. thesis defense, under the supervision of Claude Viterbo and Pierre Pansu

Daniel Perez [1,2,3]

[1]École normale supérieure

[2]Université Paris-Saclay

[3]DataShape, INRIA

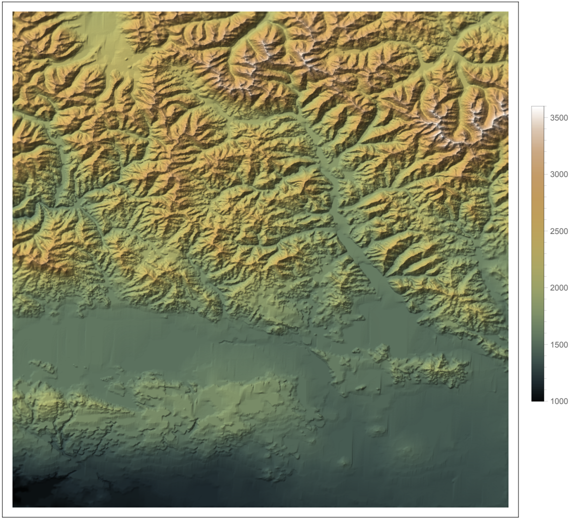
July 11th 2022

Table of contents

Introduction

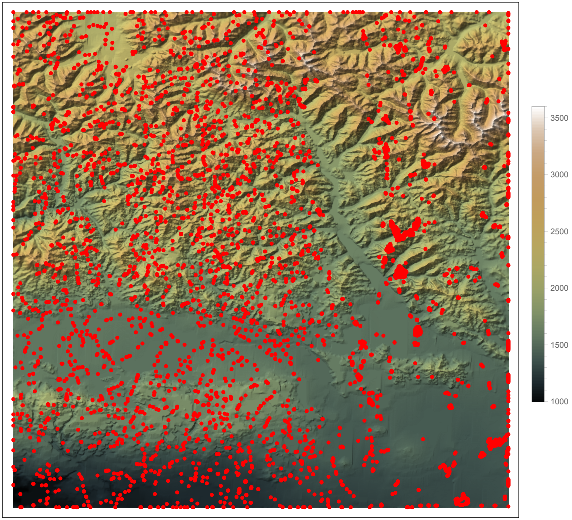
Question

How can we define a proper notion of "peak" in alpinism ?



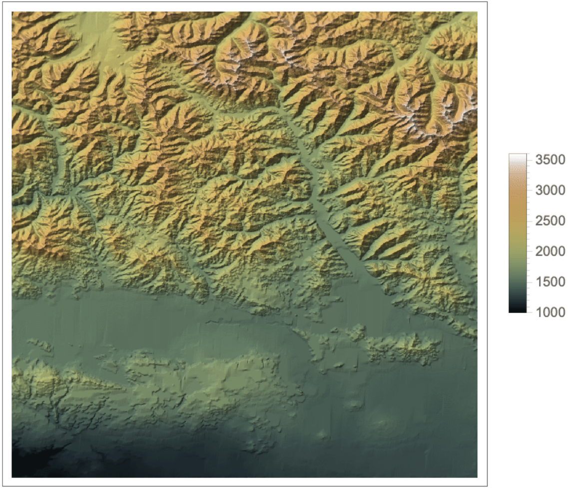
Idea

Look at the local maxima of a the elevation function of the terrain.



Idea

Flood the terrain, and look at islands forming and disappearing.



Remark

There are water levels b where an island "is born"...

FIGURE – water level = 1524 m

FIGURE – water level = 1525 m

Remark

...and water levels d where islands "die"...

FIGURE – water level = 1524 m

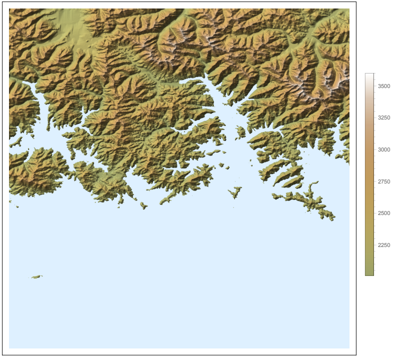


FIGURE – water level = 2005 m

Definition

We call a point a peak if the "persistence" ($:= d - b$) of its island is $\geq 91m$ (300 ft.)

We can keep track of the islands by looking at the so-called "barcode" or "diagram" of the elevation function.

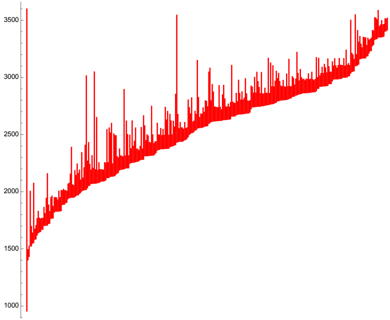


FIGURE – Barcode of the height function

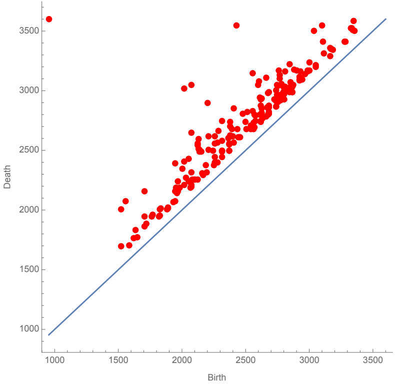


FIGURE – Diagram of the height function

Question

- *How can we easily compute this barcode object given a landscape ?*
- *Is it possible to have a computer do this task automatically ?*

Solution

Persistent homology !

- Homology $\Longrightarrow$ "counting islands" given a water level.
- Persistence $\Longrightarrow$ what happens when we change water level ?

Question

Given a space X and a function $f : X \to \mathbb{R}$ what can we say about the topology (more precisely, the homology) of the sets $\{f \geq t\}$?

A crashcourse in homology

Fact

Homology is an invariant characterizing k-dimensional holes.

Remark

Here, we will only define homology for simplicial complexes

A crashcourse in homology : simplicial complexes

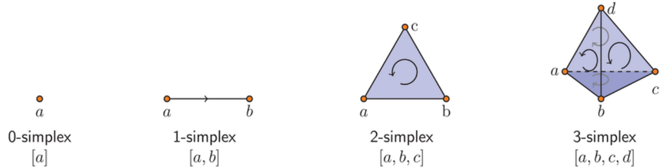


FIGURE – Oriented n-simplices

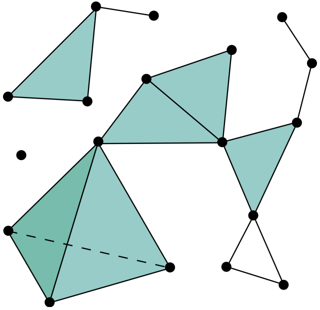


FIGURE – A simplicial complex

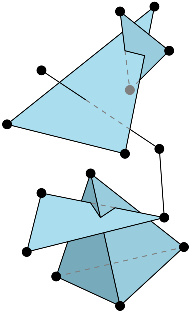


FIGURE – Not a simplicial complex

Definition

Let k be a field and X be a simplicial complex. The chain space *of degree n associated to X is*

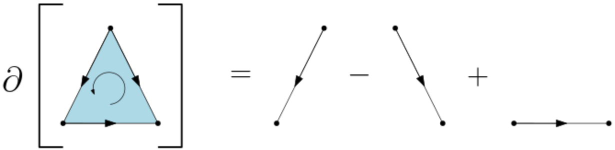
$$C_n(X) := \langle n\text{-faces of } X \rangle_k$$

A crashcourse in homology : chain complexes

Definition

The boundary operator $\partial : C_n(X) \to C_{n-1}(X)$ *is a linear operator defined by*

$$\partial(n\text{-}face) = \sum_{oriented} \textit{elements of boundary of the n-face}$$



$$\partial\left(\overset{A}{\bullet}\longrightarrow\overset{B}{\bullet}\right) = \overset{B}{\bullet} - \overset{A}{\bullet} \qquad\qquad \partial(\bullet) = 0$$

Remark

$$\partial^2 = 0$$

Definition

The couple $(C_\bullet, \partial)$ *is called a* chain complex, *it is common to note it as a sequence*

$$\cdots \xrightarrow{\partial} C_3 \xrightarrow{\partial} C_2 \xrightarrow{\partial} C_1 \xrightarrow{\partial} C_0 \xrightarrow{\partial} 0\,.$$

Remark

$$\partial(loops) = 0$$

More generally, we call an n-cycle any element of $\ker(\partial)$.

Idea

We could define an "n-hole" to be an n-cycle that is not a boundary.

Definition

The nth homology group is

$$H_n(X) := \frac{\ker(\partial|_{C_n})}{\partial C_{n+1}}$$

Here, computation of homology $\iff$ reduction of the matrix ∂.

Remark

We have just turned the problem of finding "holes" into a linear algebra problem !

Definition

The nth homology group is

$$H_n(X) := \frac{\ker(\partial|_{C_n})}{\partial C_{n+1}}$$

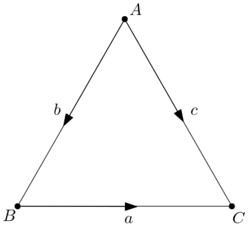
Here, computation of homology $\iff$ reduction of the matrix ∂.

Remark

We have just turned the problem of finding "holes" into a linear algebra problem !

Remark

$$H_0(X) = \frac{C_0}{\partial C_1} \approx \frac{\text{points of } X}{\text{pairs of points joined by a path}}$$
$$\approx \text{path connected components}$$

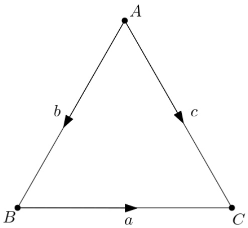


$$C_0 = \langle A, B, C \rangle_k$$
$$C_1 = \langle a, b, c \rangle_k$$
$$C_n = 0 \text{ for } n \geq 2\,.$$

The chain complex is thus

$$0 \xrightarrow{\partial} \langle a, b, c \rangle_k \xrightarrow{\partial} \langle A, B, C \rangle_k \xrightarrow{\partial} 0$$
$$a \longmapsto C - B$$
$$b \longmapsto C - A$$
$$c \longmapsto B - A$$

A crashcourse in homology : an example



$$\partial = \begin{array}{c} \\ \end{array}\begin{array}{cc} \begin{array}{cccccc} A & B & C & a & b & c \end{array} & \\ \begin{pmatrix} 0 & 0 & 0 & 0 & -1 & -1 \\ 0 & 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} & \begin{array}{c} A \\ B \\ C \\ a \\ b \\ c \end{array} \end{array} \implies \begin{cases} \ker(\partial|_{C_1}) = \langle a - b + c \rangle_k \\ \ker(\partial|_{C_0}) = \langle A, B, C \rangle_k \\ \partial C_1 = \langle C - B, C - A \rangle_k \end{cases}$$

It follows that

$$H_0(X) = \frac{\langle A, B, C \rangle_k}{\langle C - B, C - A \rangle_k} \cong \langle C \rangle_k \implies \text{1 connected component}$$

$$H_1(X) = \frac{\langle a - b + c \rangle_k}{0} \cong \langle a - b + c \rangle_k \implies \text{1 1-hole.}$$

Remark

This construction extends [6] to topological spaces homeomorphic to a simplicial complex.

Remark

Homology can be defined in any situation where we have a chain complex $(C_\bullet, \partial)$ *where* ∂ *satisfied* $\partial^2 = 0$.

Persistence ?

Now, we look at $f : X \to \mathbb{R}$ which filters the space into $(X_t)_{t\in\mathbb{R}}$ where

$$X_t = \{f \geq t\}\,.$$

Idea

$$\begin{array}{c}
\textit{Filtration by } f : X \to \mathbb{R} \\
\big\downarrow \textit{Induces} \\
\textit{Inclusions}\ \ i_{r,s} : X_r \to X_s\ (r > s) \\
\big\downarrow \textit{Induce} \\
\textit{Morphisms}\ \ C_\bullet(X_r) \to C_\bullet(X_s) \\
\big\downarrow \textit{Induce} \\
\textit{Morphisms}\ \ H_\bullet(i_{r,s}) : H_\bullet(X_r) \to H_\bullet(X_s)
\end{array}$$

To study the filtration, we need to study the morphisms $H_\bullet(i_{r,s})$.

Example

$H_0(i_{r,s})$ *tells us which islands "alive" at $t = r$ are still alive at time $t = s$.*

Definition

The persistent homology *of the couple* (X, f)*, denoted* $H_\bullet(X, f)$ *is the collection of the vector spaces* $(H_\bullet(X_t))_{t\in\mathbb{R}}$ *together with the collection of morphisms* $(H_\bullet(i_{r,s}))_{r>s}$.

We can see $H_\bullet(X, f)$ as a "map" (formally, a functor) associating :

$$t \longmapsto H_\bullet(X_t)$$

$$r > s \longmapsto H_\bullet(i_{r,s})$$

Theorem (Decomposition theorem, Auslander, Ringel, Tachikawa, Gabriel, Azumaya, Chazal, Crawley-Boevey, de Silva [1])

Under some conditions on (X, f), the persistent homology of the couple admits a decomposition

$$H_\bullet(X,f) = \bigoplus_{(b,d)} k[b,d[\,,$$

where $[b,d[\subset \mathbb{R}$ and $k[b,d[$ is the "map" (functor) associating

$$t \mapsto \begin{cases} k & t \in [b,d[\\ 0 & \textit{else} \end{cases}$$

$$r > s \mapsto \begin{cases} \mathrm{id} & r,s \in [b,d[\\ 0 & \textit{else} \end{cases}$$

Remark

The intervals $[b,d[$ are exactly the intervals we encountered earlier with the birth and death of islands !

Remark

The conditions on (X, f) will always be satisfied for the rest of this talk.

Theorem (Stability theorem, Cohen-Steiner, Edelsbrunner, Harer, Chazal, Carlsson,... [7])

Let X be a compact metric space. There exists a notion of distance on persistence diagrams or barcodes, d_∞, satisfying

$$d_\infty(\mathrm{Dgm}(f), \mathrm{Dgm}(g)) \leq \|f - g\|_\infty .$$

for any pair of continuous functions $f, g : X \to \mathbb{R}$.

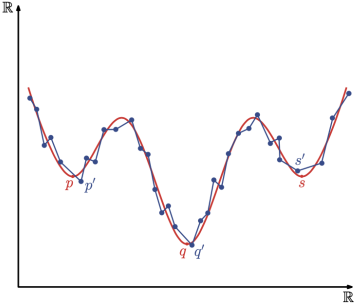


FIGURE – Two close functions

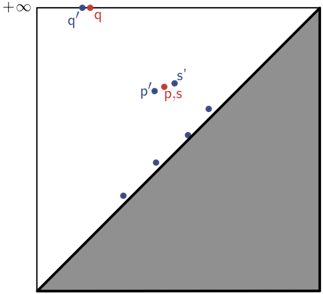


FIGURE – The diagrams of the functions are close

Why is PH interesting ?

- PH is a robust invariant ;
- Systematic approach to studying the topology of superlevel sets of a function $f : X \to \mathbb{R}$;
- Fairly general and widely applicable pipeline ;
- Lower semi-continuity of diagrams induced by the stability theorem imply non-trivial results in a wide variety of situations.

Main question

The question (partially) answered in this thesis

Question (Viterbo)

What does the barcode of a random or generic function look like ?

Question (Viterbo)

What does the barcode of a random or generic function look like ?

Solution (1)

? ? ? ?

The question (partially) answered in this thesis

Question (Viterbo)

What does the barcode of a random or generic function look like ?

Solution (1)

? ? ? ?

Solution (2)

Let us run some experiments !

Problem

What model should we fix ?

Definition

Brownian motion is the unique stochastic process B satisfying

- $B_0 = 0$ *a.s. and* $B_t \sim \mathcal{N}(0, t)$ *;*
- (Independence of increments) $\forall 0 \leq t_1 < t_2 < \cdots < t_n < \infty$, $(B_{t_{i+1}} - B_{t_i})_i$ *are mutually independent ;*
- (Stationarity of increments) $\forall s < t$, $B_t - B_s = B_{t-s}$ *in distribution.*

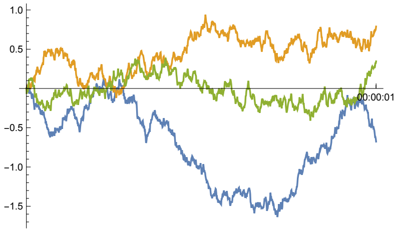


FIGURE – Three Brownian paths

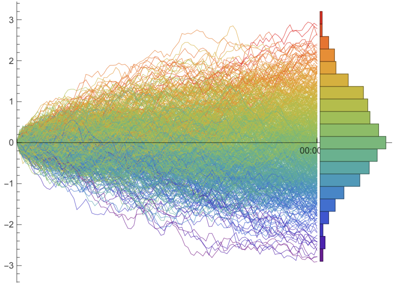


FIGURE – 750 Brownian paths, with the distribution of the values of the process at $t = 1$.

Fact

BM is very well approximated by random walks $\Longrightarrow$ easy to compute the barcode !

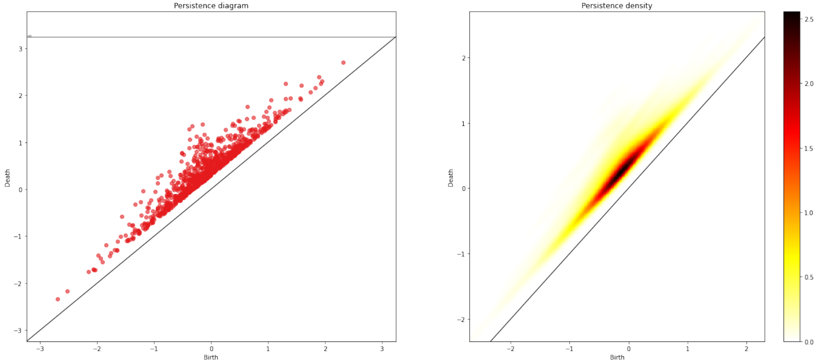


FIGURE – Average persistence diagram of 300 paths

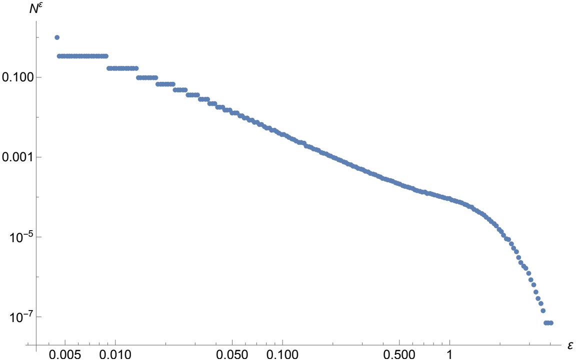


FIGURE – Number of bars of length $\geq \varepsilon$ (N^ε) of BM

Remark

There seems to be a very regular regime for small bars !

Does this work in general ?

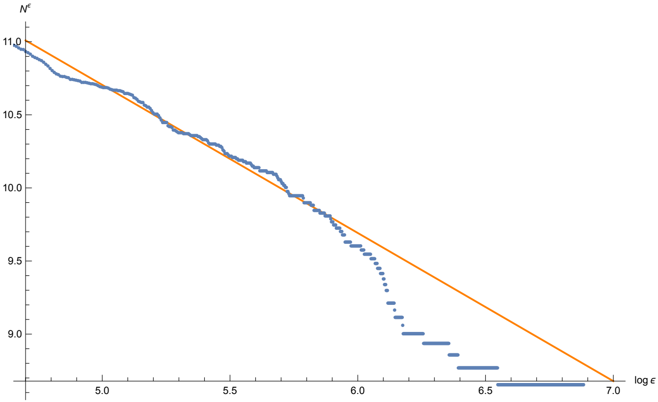


FIGURE – Number of bars of length $\geq \varepsilon$ (N^ε) of the mountain landscape

Remark

Same behaviour, but with a different value for the slope of the line !

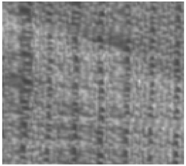
Goals

- Stability (is the approximation of BM by a process close to it accurate) ?
- Explain the behaviour of N^ε as $\varepsilon \to 0$? As $\varepsilon \to \infty$?
- Deterministic vs. random ?

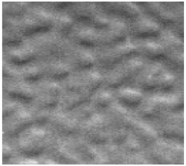
Deterministic functions

Motivation for the introduction of Wasserstein distances on diagrams

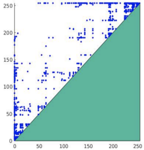
Often, crucial information is encoded in small bars !

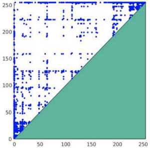


Label = Canvas



Label = Carpet





Problem

d_∞ doesn't take into account small bars ! A priori, the stability theorem is useless !

Solution

Introduce a new family of distances on diagrams.

Idea (Chazal, *et al.*, [2])

Diagrams are measures on the half-plane $\mathcal{X}$.

Idea

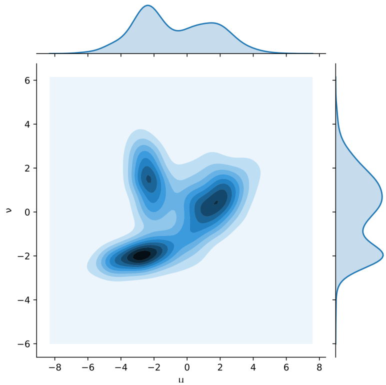
Extend this notion by introducing the vague convergence topology on the space of measures on $\mathcal{X}$ and close the space of persistence measures *with respect to this topology.*

Use optimal transport

Definition (Wasserstein distances)

Let (X, d) be a polish metric space. Denote $\mathcal{P}(X)$ the set of probability measures and $\mathcal{P}_p(X)$ the set of probability measures admitting a pth moment. The pth Wasserstein distance between $\mu, \nu \in \mathcal{P}_p(X)$,

$$W_{p,d}(\mu, \nu) := \inf_{\pi \in \Gamma(\mu,\nu)} \left[\int_{X^2} d(x,y)^p \, d\pi(x,y) \right]^{1/p} .$$



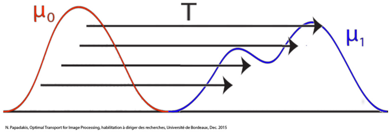


FIGURE – A transport plan

Problem

Persistence measures are not *probability measures and can have infinite mass !*

Wasserstein distances d_p

Problem

Persistence measures are not *probability measures and can have infinite mass !*

Solution (Divol, Lacombe [4])

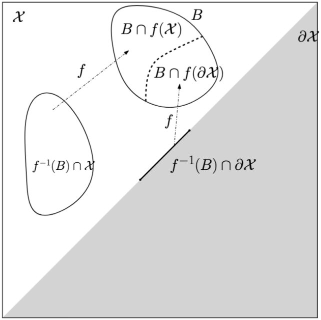
Take the diagonal to be reservoir of infinite mass, so as to be able to compare measures with different masses.

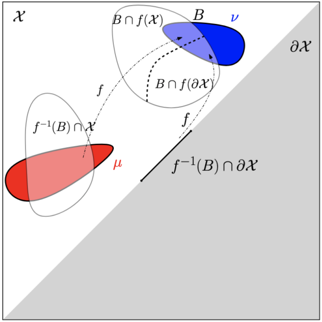
Idea

Introduce Wasserstein distances as before, this time for persistence measures.

Definition

The optimal transport distance thus obtained on the space of diagrams will be denoted d_p.





Remark

- *For* $p = \infty$, *we retrieve* d_∞ *;*
- *For* $p < \infty$, d_p *takes into account small bars.*

Remarks

Remark

- *For* $p = \infty$*, we retrieve* d_∞ *;*
- *For* $p < \infty$*,* d_p *takes into account small bars.*

Question

- *Metric spaces of diagrams, given* p *?*
- *Given a certain class of functions, which* p *should we choose to consider ?*
- *Does* d_p *admit a stability theorem with respect to perturbations in* L^∞ *?*

Divol and Lacombe [4] introduced

Definition

Denote $\mathcal{D}$ the space of Radon measures on the usual half-plane, then

$$\mathcal{D}_p := \{D \in \mathcal{D} \mid d_p(D, \Delta) < \infty\}$$

Motivates considering

Definition

$$\mathrm{Pers}_p(D) := d_p(D, \Delta)\,.$$

Divol and Lacombe [4] introduced

Definition

Denote $\mathcal{D}$ the space of Radon measures on the usual half-plane, then

$$\mathcal{D}_p := \{D \in \mathcal{D} \,|\, d_p(D, \Delta) < \infty\}$$

Motivates considering

Definition

$$\mathrm{Pers}_p(D) := d_p(D, \Delta)\,.$$

Remark

For a persistence diagram D,

$$\mathrm{Pers}_p^p(D) = \sum_{b \in D} \ell(b)^p$$

where $\ell(b)$ is the length of the bar corresponding to the point $b \in D$.

Definition

The kth persistence index *is defined as*

$$\mathcal{L}_k(f) := \inf\{p \geq 1 \,|\, \mathrm{Pers}_p(H_k(X, f)) < \infty\}$$

Associating the correct metric space to $\mathrm{Dgm}(f)$

Theorem (Picard, §3 [10])

Let $f : [0,1] \to \mathbb{R}$ *be a continuous function and denote*

$$\mathcal{V}(f) := \inf\{p \mid \|f\|_{p\text{-}var} < \infty\}\,.$$

Then,

$$\mathcal{V}(f) = \mathcal{L}_0(f) = \limsup_{\varepsilon \to 0} \frac{\log N^\varepsilon}{\log(1/\varepsilon)} \vee 1$$

where $a \vee b := \max\{a, b\}$.

Remark

This theorem already entails that for BM, since $\mathcal{V}(B) = 2$, *we have* $\forall \delta > 0$,

$$N^\varepsilon = O(\varepsilon^{-2-\delta}) \text{ as } \varepsilon \to 0\,.$$

A digression

Theorem (Picard, [10])

Due to self-similarity properties of BM, N^ε is a.s.

$$N^\varepsilon \sim \frac{C}{\varepsilon^2} \text{ as } \varepsilon \to 0\,.$$

for a certain constant C. Furthermore,

$$\mathbb{E}[N^\varepsilon] = \frac{C}{\varepsilon^2} + O(1) \text{ as } \varepsilon \to 0\,.$$

Question

What about large ε ? Does this asymptotic expansion extend ?

Theorem (Persistence-Regularity theorem, P., [8])

Let X be a d-dimensional compact Riemannian manifold, $k \in \mathbb{N}$ and let $f \in C^\alpha(X, \mathbb{R})$, then

$$\mathcal{L}_k(f) = \limsup_{\varepsilon \to 0} \frac{\log(N_k^\varepsilon)}{\log(1/\varepsilon)} \vee 1 \leq \frac{d}{\alpha}\,.$$

Remark

$$\mathcal{L}_k(f) = q \iff \forall \delta > 0,\ N_k^\varepsilon = O(\varepsilon^{-q-\delta}) \text{ as } \varepsilon \to 0\,.$$

Theorem (Persistence-Regularity theorem, P., [8])

Let X be a d-dimensional compact Riemannian manifold, $k \in \mathbb{N}$ and let $f \in C^\alpha(X, \mathbb{R})$, then

$$\mathcal{L}_k(f) = \limsup_{\varepsilon \to 0} \frac{\log(N_k^\varepsilon)}{\log(1/\varepsilon)} \vee 1 \leq \frac{d}{\alpha} .$$

Remark

$$\mathcal{L}_k(f) = q \iff \forall \delta > 0,\ N_k^\varepsilon = O(\varepsilon^{-q-\delta}) \text{ as } \varepsilon \to 0 .$$

Inspiring ourselves from a previous work of Baryshnikov and Weinberger,

Theorem (Genericity theorem, P., [8])

The inequality above is saturated generically (in the sense of Baire) in $C^\alpha(X, \mathbb{R})$ for $k \in \{0, \cdots, d-1\}$.

Theorem (Persistence-Regularity theorem, P., [8])

Let X be a d-dimensional compact Riemannian manifold, $k \in \mathbb{N}$ and let $f \in C^\alpha(X, \mathbb{R})$, then

$$\mathcal{L}_k(f) = \limsup_{\varepsilon \to 0} \frac{\log(N_k^\varepsilon)}{\log(1/\varepsilon)} \vee 1 \leq \frac{d}{\alpha}\,.$$

Remark

$$\mathcal{L}_k(f) = q \iff \forall \delta > 0,\ N_k^\varepsilon = O(\varepsilon^{-q-\delta}) \text{ as } \varepsilon \to 0\,.$$

Inspiring ourselves from a previous work of Baryshnikov and Weinberger,

Theorem (Genericity theorem, P., [8])

The inequality above is saturated generically (in the sense of Baire) in $C^\alpha(X, \mathbb{R})$ for $k \in \{0, \cdots, d-1\}$.

Corollary

$$f \in C^\alpha(X, \mathbb{R}) \implies \mathrm{Dgm}(f) \in \mathcal{D}_p \text{ for } p > \frac{d}{\alpha}\,,$$

but not in $\mathcal{D}_p$ for $p < \frac{d}{\alpha}$ generically.

Wasserstein stability revisited

Theorem (Wasserstein stability, P. [8])

Let X be a d-dimensional compact Riemannian manifold and $f, g \in C^{\alpha}_{\Lambda}(X, \mathbb{R})$. Then, for every $k \in \mathbb{N}^*$, and all $p > q > \frac{d}{\alpha}$,

$$d_p(H_k(X, f), H_k(X, g)) \leq C_{X,\Lambda,\alpha,q,k} \, \|f - g\|_{\infty}^{1-\frac{q}{p}} .$$

Remark

We have explicit estimates on C_X in terms of metric quantities of X and under all the hypotheses the bounds on q are sharp.

Remark

Genericity theorem $\implies$ for $p < \frac{d}{\alpha}$, WS generically violated in $C^{\alpha}(X, \mathbb{R})$.

- Wasserstein stability was previously shown for Lipschitz functions, without quantitative conditions on C_X and q by Cohen-Steiner, Edelsbrunner, Harer and Mileyko [3] in 2010 ;
- Skraba and Turner [11] showed in 2020 that Cohen-Steiner *et al.*'s stability theorem satisfied $\implies q \geq d$;
- Our result on genericity shows the other direction : $q \geq d \implies$ stability

Take-away message

- Persistence index $\iff$ Asymp. behaviour of N^ε as $\varepsilon \to 0$ $\iff$ regularity of f ;
- Quantitative stability results which take into account small bars exist, under some hypotheses on regularity ;
- The "worst-case" scenario for $\mathcal{L}_k(f)$ is generically attained $\implies$ range of p for d_p which should be considered in applications.

Stochastic processes

A notion of an average diagram

Definition

From now on, f and g will always be two $\mathbb{R}$-valued a.s. C^α stochastic processes defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on a d-dimensional compact Riemannian manifold X.

Definition

This notion of average diagram is given by dual action as a Radon measure, i.e. for every measurable set B of the half-space, we define

$$\mathbb{E}[\mathrm{Dgm}(f)]\,(B) := \mathbb{E}[\mathrm{Dgm}(f)(B)]\ .$$

Proposition (P., [8])

- *$H_k(X, f) \in \mathcal{D}_r$ for every $\frac{d}{\alpha} < r < \infty$.*
- *If $\frac{d}{\alpha} < q < \infty$ and $\mathbb{E}\left[\|f\|^q_{C^\alpha(X,\mathbb{R})}\right] < \infty$, then $\mathbb{E}[H_k(X, f)] \in \bigcap_{\frac{d}{\alpha} < p \leq q} \mathcal{D}_p$.*

Declinations of stability

Theorem (Stochastic Wasserstein stability, P. [8])

If the supports of $f_\sharp\mathbb{P}$ and $g_\sharp\mathbb{P}$ are compact, then for every $\frac{d}{\alpha} < p \leq q \leq \infty$ there exists a constant $C_{X,p,\eta}$ such that

$$d_p(\mathbb{E}[H_k(X,f)]\,,\mathbb{E}[H_k(X,g)]) \leq C_{X,p,\eta} W^{\eta}_{q\eta,L^\infty}(f_\sharp\mathbb{P}, g_\sharp\mathbb{P})\,.$$

where $\eta < 1 - \frac{d}{\alpha p}$.

By using the Komlós–Major–Tusnády approximation

Corollary

In the case of BM being approximated by a random walk W_n with n steps, for $p > 2$,

$$d_p(\mathbb{E}[\mathrm{Dgm}(B)]\,,\mathbb{E}[\mathrm{Dgm}(W_n)]) \leq O(n^{-\frac{1}{2}}\log(n))\,.$$

Zeta functions associated to a stochastic process

Definition

The ζ-function in degree k associated to a stochastic process f, denoted $\zeta_{f,k}$ is defined as

$$\zeta_{f,k}(p) := \mathbb{E}\left[\mathrm{Pers}_p^p(H_k(X,f))\right] .$$

We denote $\zeta_{f,0} =: \zeta_f$.

Remark

This is a priori well-defined only on a strip of the complex plane.

Proposition (P., [9])

N^ε and Pers_p^p are dual, in the sense that

$$\zeta_f(p) = p \int_0^\infty \varepsilon^{p-1} \mathbb{E}[N^\varepsilon] \; d\varepsilon \tag{1}$$

$$\mathbb{E}[N^\varepsilon] = \frac{1}{2\pi i} \int_{\gamma - i\infty}^{\gamma + i\infty} \varepsilon^{-p} \zeta_f(p) \; \frac{dp}{p} \,. \tag{2}$$

Fact (Flajolet [5])

The following correspondence holds

$$\text{Meromorphic ext. of } \zeta \text{ (to the left, to the right)}$$
$$\Longleftrightarrow$$
$$\text{Asymptotic developments of } \mathbb{E}[N^\varepsilon] \,(\text{as } \varepsilon \to 0, \varepsilon \to \infty)\,.$$

Remark

This applies in general to functions and their Mellin transforms.

Renewal theory allows us to calculate

Theorem (P., [9])

The ζ-function of Brownian motion on the interval $[0, t]$ admits a meromorphic extension to the whole complex plane. Furthermore, it is exactly equal to

$$\zeta_B(p) = \frac{4(2^p - 3)}{\sqrt{\pi}} \left(\frac{t}{2}\right)^{\frac{p}{2}} \Gamma\left(\frac{p+1}{2}\right) \zeta(p-1)$$

for all p and has a unique simple pole at $p = 2$ of residue $[B]_t = t$.

From which we deduce

Proposition (P., [9])

For Brownian motion on $[0, t]$

$$\begin{aligned}\mathbb{E}[N^\varepsilon] &= 4 \sum_{k \geq 1} (2k-1)\,\mathrm{erfc}\left(\frac{(2k-1)\varepsilon}{\sqrt{2t}}\right) - k\ \mathrm{erfc}\left(\frac{2k\varepsilon}{\sqrt{2t}}\right) \\ &= \frac{t}{2\varepsilon^2} + \frac{2}{3} + 2 \sum_{k \geq 1} (2(-1)^k - 1) \frac{e^{-\pi^2 k^2 t / 2\varepsilon^2} t}{\varepsilon^2} \left[1 + \frac{\varepsilon^2}{\pi^2 k^2 t}\right].\end{aligned}$$

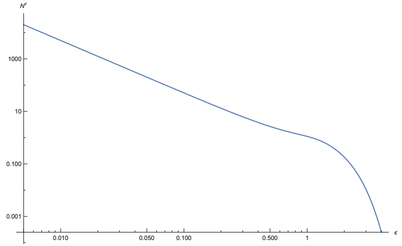


FIGURE – Theoretical prediction of N^ε of BM

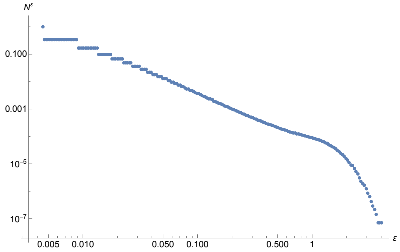


FIGURE – Experimental results

Remark

For Brownian motion, we also have an analytic expression for the density of $\mathbb{E}[H_0([0,t],B)]$ and even the distribution of the length of the kth longest bar.

Definition

A Lévy process is a stochastic process f satisfying

- $f(0) = 0$ *a.s. ;*
- (Independence of increments) $\forall 0 \leq t_1 < t_2 < \cdots < t_n < \infty$, $(f(t_{i+1}) - f(t_i))_i$ *are mutually independent ;*
- (Stationarity of increments) $\forall s < t$, $f(t) - f(s) = f(t-s)$ *in distribution.*
- (Continuity in probability) $\forall \varepsilon > 0, t \geq 0$,

$$\lim_{h \to 0} \mathbb{P}(|f(t+h) - f(t)| > \varepsilon) = 0\,.$$

Definition

f is said to be α-stable if for all t $f(\lambda^\alpha t) = \lambda f(t)$ in distribution.

Lévy processes

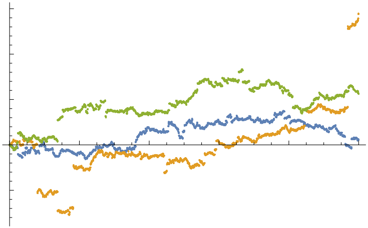


Figure – Three 1.5-stable paths

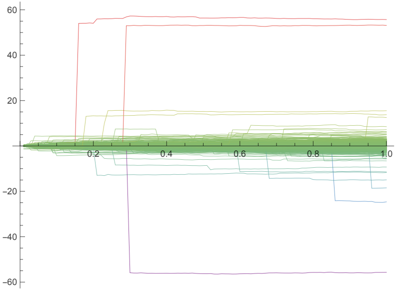


Figure – 750 1.5-stable paths

Using the same techniques,

Theorem (P., [9])

Let f be an α-stable Lévy process on $[0, t]$, then for some constants A and B,

$$\mathbb{E}[N^\varepsilon] = \frac{t}{A\varepsilon^\alpha} + B + o(\varepsilon^{\alpha n}) \quad \text{as } \varepsilon \to 0\,.$$

Remark

Brownian motion is a 2-stable Lévy process.

Stochastic stability $\implies$ (Different avg diagrams $\implies$ Different distributions)

Remark

Can use this to discern stochastic processes !

Statistical test

This a parameter test for α, given an α-stable process on $[0,1]$.

1. *Sample M paths of the stochastic process f (for example at regular intervals of size $\frac{1}{N}$ for some N) ;*
2. *Compute the barcode of the sampled paths.*
3. *For some range of small enough ε, and for some positive constant $c > 1$ compute the quantity*

$$\hat{\alpha}_M := \log_c \left[\frac{\overline{N}_t^{\varepsilon/c} - \overline{N}_t^{2\varepsilon/c}}{\overline{N}_t^{\varepsilon} - \overline{N}_t^{2\varepsilon}} \right].$$

Claim

$\hat{\alpha}_M$ is a good estimation of α.

Parameter testing

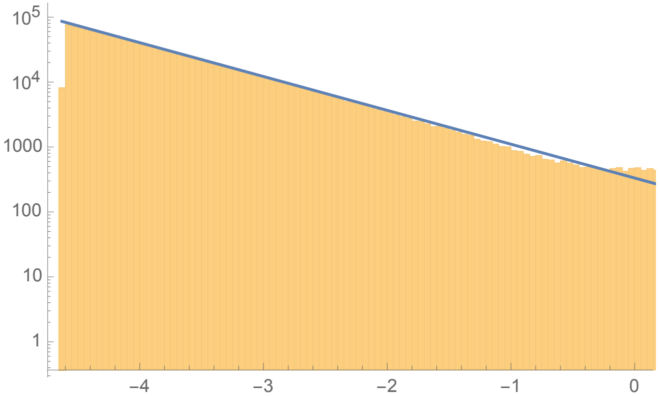


FIGURE – In orange, a histogram of the number of bars of length $\geq \varepsilon$, N^ε found as a function of $\log \varepsilon$ from a simulation of a Lévy 1.2-stable process as a random walk. In blue, the function $C_{1.2}\varepsilon^{-1.2}$.

Conclusion

- PH is a good and robust invariant to study stochastic processes ;
- Persistence indices are related to the asymptotic behaviour of N^{ε} , which is itself related to self-similarity and regularity of the process ;
- The genericity results regarding the link between regularity and the persistence index gives an appropriate range of p which should be considered in applications ;
- By stochastic stability, it is possible to create new statistical tests, able to discern different statistical processes.

- Further explore the connection between regularity and the persistence index for processes of higher regularity than C^α ;
- Up to redefinition of the persistence indices, it may be possible to extend the theory to some spaces of infinite dimension ;
- Applications : spin glasses (?), machine learning, symplectic geometry,...

F. Chazal, W. Crawley-Boevey, and V. de Silva.
The observable structure of persistence modules.
Homology, Homotopy and Applications, 18(2) :247–265, 2016.

F. Chazal, V. de Silva, M. Glisse, and S. Oudot.
The Structure and Stability of Persistence Modules.
Springer International Publishing, 2016.

D. Cohen-Steiner, H. Edelsbrunner, J. Harer, and Y. Mileyko.
Lipschitz functions have L^p-stable persistence.
Foundations of Computational Mathematics, 10(2) :127–139, Jan 2010.

V. Divol and T. Lacombe.
Understanding the topology and the geometry of the persistence diagram space via optimal partial transport.
CoRR, abs/1901.03048, 2019.

P. Flajolet, X. Gourdon, and P. Dumas.
Mellin transforms and asymptotics : Harmonic sums.
Theoretical Computer Science, 144(1) :3–58, 1995.

A. Hatcher.
Algebraic topology.
Cambridge University Press, Cambridge, 2002.

S. Y. Oudot.
Persistence Theory - From Quiver Representations to Data Analysis, volume 209 of *Mathematical surveys and monographs*.
American Mathematical Society, 2015.

D. Perez.
On C^0-persistent homology and trees.
https://arxiv.org/abs/2012.02634, Dec. 2020.

D. Perez.
ζ-functions and the topology of superlevel sets of stochastic processes.
arXiv e-prints, page arXiv :2110.10982, Oct. 2021.

J. Picard.
A tree approach to p-variation and to integration.
The Annals of Probability, 36(6) :2235–2279, Nov 2008.

P. Skraba and K. Turner.
Wasserstein stability for persistence diagrams, 2020.