Algebraic Tools for TDA in a Multiparameter Setting

based on joint work with Wojciech Chachólski and René Corbet

arXiv:2112.06509

Anna-Laura Sattelberger (MPI-MiS Leipzig)

AlToGeLiS seminar
February 24, 2022

MAX-PLANCK-GESELLSCHAFT

Topological Data Analysis

Studying the shape of data...

...with tools from (algebraic) topology

Applications

- ◇ medical and life sciences
- ◇ distinguishing point processes on a unit square (Chachólski–Riihimäki, 2020)
- ◇ topological machine learning

...whenever data arise!

Behind the scenes

- ◇ commutative algebra
- ◇ algebraic geometry

New invariants of multigraded modules...

...arising from TDA

Barcoding

Main tool: persistent homology

Associating barcodes to data

Input: point cloud $\{p_i\} \subseteq \mathbb{R}^N$

1. $X_\varepsilon := \cup_{p_i} B_\varepsilon(p_i)$
2. increase $\varepsilon \overset{\text{nerve}}{\rightsquigarrow}$ filtered simplicial complex
3. for all n: n-th homology with coefficients in $\mathbb{K}$ naturally is a finitely generated $\mathbb{N}$-graded module P_n over $\mathbb{K}[t]$
4. structure theorem for finitely generated modules over PIDs:

$$P_n \cong \bigoplus_i \mathbb{K}[t]t^{\alpha_i} \oplus \bigoplus_j \mathbb{K}[t]t^{\beta_j}/\mathbb{K}[t]t^{\beta_j+\gamma_j}$$

Output: barcode $\{[\alpha_i, \infty), [\beta_j, \beta_j + \gamma_j)\}$

Fact: This invariant is discrete, complete, and stable.

Multiparameter persistence

Study of **multifiltered** simplicial complexes (Carlsson–Zomorodian, 2009)

Algebraic counterpart

$\mathbb{N}^r$-graded $\mathbb{K}[x_1, \ldots, x_r]$-modules $M = \oplus_{a \in \mathbb{N}^r} M_a$

Challenges

- no higher-dimensional analogue of barcodes
- lack of **stable**, **algorithmic** invariants

Multipersistence modules as functors

Let $G \in \{\mathbb{N}^r, \mathbb{R}^r_{\geq 0}\}$ (more general monoids in (Corbet–Kerber, 2018))

$$\begin{array}{ccc} \mathrm{Fun}((G,\leq),\mathrm{Vect}_{\mathbb{K}}) & \overset{\text{isom. of cats}}{\cong} & G\text{-graded } \mathbb{K}[G]\text{-modules} \\ \cup & & \cup \\ \mathrm{Tame}((G,\leq),\mathrm{Vect}_{\mathbb{K}}) & \cong & \text{finitely presented } G\text{-graded } \mathbb{K}[G]\text{-modules} \end{array}$$

Turning discrete into stable invariants

T	a set
f	a discrete invariant $f\colon T \to \mathbb{N}$
d	an extended pseudometric $d\colon T \times T \to \mathbb{R}_{\geq 0} \cup \{\infty\}$
$\mathcal{M}$	measurable functions $[0,\infty) \to [0,\infty)$ endowed with interleaving distance

Definition & Theorem (Gäfvert–Chachólski, 2017)

The **hierarchical stabilization** of f at $x \in T$, denoted $\widehat{f}(x) \in \mathcal{M}$, is

$$\widehat{f}(x)(\tau) \coloneqq \min\{f(y) \mid y \in T\colon d(x,y) \leq \tau\}.$$

For any choice of d, $\widehat{f}\colon T \to \mathcal{M}$ is 1-Lipschitz.

Measuring distances between tame functors

How to construct metrics—in best case in a way that is suitable for learning tasks?

Pseudometrics arising from contours

$\mathbf{R}^r_\infty$ the poset obtained from adding one element ∞ to $\mathbb{R}^r_{\geq 0}$

Definition

A **persistence contour** is a functor $C\colon \mathbf{R}^r_\infty \times \mathbb{R}_{\geq 0} \to \mathbf{R}^r_\infty$ such that for every $x \in \mathbf{R}^r_\infty$, $\tau, \varepsilon \in \mathbb{R}_{\geq 0}$:

1. $x \leq C(x, \varepsilon)$ and
2. $C(C(x, \varepsilon), \tau) \leq C(x, \varepsilon + \tau)$.

Example (Standard contour)

$C(x, \varepsilon) = x + \varepsilon \cdot v$, where $v = (v_1, \ldots, v_r) \in \mathbb{R}^r_{\geq 0}$

ε-neighborhoods of 0

For $\varepsilon \in \mathbb{R}_{\geq 0}$ define

$$\mathcal{D}_\varepsilon := \{G \in \mathrm{Tame}(\mathbb{R}^r_{\geq 0}, \mathrm{Vect}_{\mathbb{K}}) \mid C(x, \varepsilon) \neq \infty \Rightarrow G(x \leq C(x, \varepsilon)) = 0\}.$$

Multigraded Betti numbers

Hilbert's syzygy theorem

Every f.g. $\mathbb{N}^r$-graded $\mathbb{K}[x_1, \ldots, x_r]$-module M has a minimal free resolution $F_\bullet$ of length at most r, i.e., there exists an exact sequence of $\mathbb{N}^r$-graded modules

$$F_\bullet : \; F_r \xrightarrow{\delta_r} \cdots \longrightarrow F_0 \xrightarrow{\delta_0} M \longrightarrow 0,$$

such that the ranks of the F_i are simultaneously minimized.

Definition

The rank of F_i in a minimal free resolution of M as above is called the **i-th total multigraded Betti number** of M and is denoted by $\beta_i(M)$.

Computing $\widehat{\beta_0}$ . . .

- . . . is NP-hard in general (Gäfvert–Chachólski, 2017)
- linear-time algorithm for quotients of monomial ideals in the bivariate case (Chachólski–Corbet–S., 2021)

A new invariant of multigraded modules

M a finitely generated $\mathbb{N}^r$-graded $\mathbb{K}[x_1, \ldots, x_r]$-module

Theorem & Definition (Chachólski–Corbet–S., 2021)

The hierarchical stabilization of β_0 w.r.t. the metric arising from the standard contour in the direction of $v = (v_1, \ldots, v_r) \in \mathbb{N}^r$ gives rise to

$$\dim_v(M) = \min\{\, \ell \mid \exists m_1, \ldots, m_\ell \in M \colon\ x_1^{v_1} \cdots x_r^{v_r} \cdot M \subseteq \langle m_1, \ldots, m_\ell \rangle \},$$

the **shift-dimension** of M. Such $\{m_1, \ldots, m_\ell\}$ **v-generate** M and are a **v-basis** of M for $\ell = \dim_v(M)$.

To be, or not to be in a v-basis, that is the question.

An element $m \in M$ can be extended to a v-basis of M if and only if

$$\dim_v(M/\langle m \rangle) = \dim_v(M) - 1.$$

Visualization

$M = \langle x_1^3 x_2, x_1 x_2^3 \rangle / \langle x_1^4 x_2^4 \rangle \in \mathrm{Mod}(\mathbb{K}[x_1, x_2])$, $v = (1,1) \in \mathbb{N}^2$

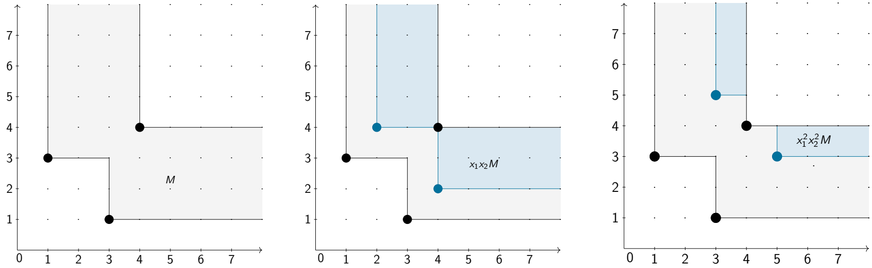

Figure: Visualization of M, $x_1 x_2 M$, and $x_1^2 x_2^2 M$

One reads:

- ◇ M, $x_1 x_2 M \subseteq \langle x_1^3 x_2, x_1 x_2^3 \rangle$, $x_1^2 x_2^2 M \subseteq \langle x_1^3 x_2^3 \rangle$, and $x_1^3 x_2^3 M = 0$.
- ◇ $\dim_{(0,0)}(M) = \dim_{(1,1)}(M) = 2$, $\dim_{(2,2)}(M) = 1$, and $\dim_{(3,3)}(M) = 0$.

Algebraic properties of the shift-dimension

Epimorphisms

If $\varphi\colon M \twoheadrightarrow N$, then $\dim_v(M) \geq \dim_v(N)$.

Proof: The image of a v-basis of M v-generates N.

Short exact sequences

Let $0 \longrightarrow M \longrightarrow L \longrightarrow N \longrightarrow 0$ be a short exact sequence of persistence modules. Then for all $v, w \in \mathbb{N}^r$, the following two inequalities hold:

1. $\dim_{v+w}(L) \leq \dim_v(M) + \dim_w(N)$,
2. $\dim_v(L) \leq \dim_v(N) + \beta_0(M)$.

Non-additivity

In general, $\dim_v(M \oplus N) \neq \dim_v(M) + \dim_v(N)$.

Last but not least

Outlook to future work

- ◇ extension of our algorithm
- ◇ application to data: which information does the shift-dimension reveal?
- ◇ construction of further multipersistence contours
- ◇ stabilization of invariants other than β_0

Thank your very much for your attention!

References I

[Bau21] Ulrich Bauer.
`Ripser`: efficient computation of Vietoris–Rips persistence barcodes.
J. Appl. Comp. Topol., 2021.

[Bub15] Peter Bubenik.
Statistical topology data analysis using persistence landscapes.
J. Mach. Learn. Res., 16:77–102, 2015.

[CCI+20] Matthieu Carrière, Frédéric Chazal, Yuichi Ike, Théo Lacombe, Martin Royer, and Yuhei Umeda.
`PersLay`: A neural network layer for persistence diagrams and new graph topological signatures.
In *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics*, volume 108 of *PMLR*, pages 2786–2796, 2020.

[CCS21] Wojciech Chachólski, René Corbet, and Anna-Laura Sattelberger.
The shift-dimension of multipersistence modules.
Preprint arXiv:2112.06509, 2021.

[CK18] René Corbet and Michael Kerber.
The representation theorem of persistence revisited and generalized.
J. Appl. and Comput. Topology, 2:1–31, 2018.

[CR20] Wojciech Chachólski and Henri Riihimäki.
Metrics and stabilization in one parameter persistence.
SIAM J. Appl. Algebra Geom., 4(1):69–98, 2020.

[CSEH07] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer.
Stability of persistence diagrams.
Discrete Comput. Geom., 37:103–120, 2007.

[CSV17] Wojciech Chachólski, Martina Scolamiero, and Francesco Vaccarino.
Combinatorial presentation of multidimensional persistent homology.
J. Pure Appl. Algebra, 221:1055–1075, 2017.

[CZ09] Gunnar Carlsson and Afra Zomorodian.
The Theory of Multidimensional Persistence.
Disc. Comp. Geom., 42:71–93, 2009.

References II

[GC17] Oliver Gäfvert and Wojciech Chachólski.
Stable invariants for multidimensional persistence.
arXiv:1703.03632, 2017.

[Ghr08] Robert Ghrist.
Barcodes: the persistent topology of data.
Bull. Amer. Math. Soc., 45:61–75, 2008.

[HMR21] Felix Hensel, Michael Moor, and Bastian Rieck.
A survey of topological machine learning methods.
Front. Artif. Intell., 4, 2021.

[MS05] Ezra Miller and Bernd Sturmfels.
Combinatorial Commutative Algebra, volume 227 of *Graduate Texts in Mathematics*.
Springer-Verlag, New York, 2005.

[RCB21] Raphael Reinauer, Matteo Caorsi, and Nicolas Berkouk.
`Persformer`: A transformer architecture for topological machine learning.
Preprint arXiv:2112.15210, 2021.

[SCL+17] Martina Scolamiero, Wojciech Chachólski, Anders Lundman, Ryan Ramanujam, and Sebastian Öberg.
Multidimensional persistence and noise.
Found. Comput. Math, 17(6):1367–1406, 2017.

[VBM+21] Oliver Vipond, Joshua A. Bull, Philip S. Macklin, Ulrike Tillmann, Christopher W. Pugh, Helen M. Byrne, and Heather A. Harrington.
Multiparameter persistent homology landscapes identify immune cell spatial patterns in tumors.
Proc. Natl. Acad. Sci. USA, 118(41), 2021.

[Vip20] Oliver Vipond.
Multiparameter persistence landscapes.
J. Mach. Learn. Res., 21:1–38, 2020.

[ZC05] Afra Zomorodian and Gunnar Carlsson.
Computing persistent homology.
Discrete Comput. Geom., 33(2):249–274, 2005.