Persistence Spheres: Bi-Continuous Linear Representations of Persistence Diagrams. Some Early Stage Results.

Matteo Pegoraro matteope@kth.se
Department of Mathematics, KTH, Stockholm

Editors: Michael Bleher, Freya Jensen, Levin Maier, Diaaeldin Taha, and Anna Wienhard

Abstract

In this extended abstract, we present ongoing work on a novel functional representation of persistence diagrams (PDs). Building on the approach of Gotovac Dogaš and Mandarić (2025), we model PDs as scalar fields on the sphere using the lift zonoid representation of finite integrable measures. Unlike their method, however, our construction yields a bi-continuous operator that is stable with respect to the 1-Wasserstein distance.

1. Introduction

Topological Data Analysis (TDA) is a rapidly growing field that applies concepts from algebraic topology to study the shape of data, providing robust, coordinate-free tools for extracting meaningful patterns, especially in complex, high-dimensional, or noisy settings.

Central to TDA is the idea that persistent homology, which is a fundamental tool that captures the multi-scale topological features of a dataset. By tracking how features such as connected components, loops, and voids appear and disappear across different spatial or scale parameters, persistent homology provides a compact summary of the data's shape. This information is typically encoded in persistence diagrams (PDs) or barcodes, which offer stable, interpretable representations well suited for both qualitative insights and some kind of quantitative analysis. In particular, each point in a persistence diagram corresponds to a feature, such as a connected component, loop, or void, recording the scale at which it appears (birth) and disappears (death). See (Edelsbrunner and Harer, 2010; Oudot, 2015) for a detailed introduction.

Data Analysis with Persistence Diagrams. To enable data analysis pipelines that leverage topological information, PDs are often compared using Wasserstein distances based on partial optimal transport (POT) (Divol and Lacombe, 2021). While these metrics are essential for ensuring robustness, they endow PDs with a non-linear geometry, which limits the range of statistical tools that can be effectively applied. Even basic operations such as averaging, typically approached via Wasserstein barycenters (Mileyko et al., 2011), are computationally challenging and may yield non-unique solutions.

Vectorizations and Kernel Methods. To address these challenges, a wide range of vectorization techniques have been developed to embed PDs into linear spaces, enabling the use of classical machine learning and statistical methods. For a comprehensive overview,

© 2025 M. Pegoraro.

we refer the reader to Pun et al. (2022); Ali et al. (2023); here, we provide only a high-level summary. First, we distinguish between methods which give an explicit embedding into a linear space, from kernel methods (Reininghaus et al., 2015; Kusano et al., 2018; Carriere et al., 2017), which, instead, rely on the kernel trick to implicitly define the feature map. The first class of methods, in turn, can be subdivided into methods characterizing diagrams via a vector of descriptive statistics (Asaad et al., 2022), algebraic methods relying on polynomial rings and tropical coordinates (Kališnik, 2019; Monod et al., 2019; Di Fabio and Ferri, 2015), and functional methods which assign to a diagram a scalar field over some domain (Bubenik, 2015; Adams et al., 2017; Biscio and Møller, 2019; Gotovac Dogaš and Mandarić, 2025).

Main Contributions In this preliminary work, we modify the approach of Gotovac Dogaš and Mandarić (2025) and propose a functional representation that maps a persistence diagram D to a scalar field $\varphi : \mathbb{S}^2 \to \mathbb{R}$ via lift zonoids. Our full version (Pegoraro, 2025) proves that this operator is Lipschitz with respect to the 1-Wasserstein distance and that, wherever it is defined, the inverse map is continuous. To the best of our knowledge, no other functional representation of PDs enjoys comparable stability and inverse-continuity guarantees. For proofs, additional details, and further experiments, we refer the reader to Pegoraro (2025).

2. Preliminaries

2.1. Convex Sets and Support Functions

We need to use the following pieces of notation, which can be found on all classical books on convex analysis/geometry. See, for instance, (Rockafellar, 1997; Salinetti and Wets, 1979). We also report the actual definitions in Section A.

Given two convex subsets $A, B \subset \mathbb{R}^3$, we indicate with $A \oplus B$ their Minkowski sum and with λA the multiplication of A with a non-negative scalar. For a compact convex set $A \subset \mathbb{R}^3$, we indicate with $h_A : \mathbb{R}^3 \to \mathbb{R}$ its support function, and for two compact subsets $A, B \subset \mathbb{R}^3$, $d_H(A, B)$ is their Hausdorff distance.

We recall that 1) any support function h_A is completely determined by its restriction on $\mathbb{S}^2$; 2) the operator $A \mapsto h_A$ is linear w.r.t. the operations between convex sets defined above:

$$\lambda_1 A \oplus \lambda_2 B \mapsto \lambda_1 h_A + \lambda_2 h_B.$$

We also need the following classical result.

Proposition 1 *Given two compact convex sets $A, B \subset \mathbb{R}^3$, the following holds:* $\max_{v \in \mathbb{S}^2} |h_A(v) - h_B(v)| = d_H(A, B)$. *In particular, the operator $A \mapsto h_A$ is injective.*

2.2. Persistence Diagrams and Measures

We define the following pieces of notation:

$$\mathbb{R}^2_{x<y} := \{(x, y) \in \mathbb{R}^2 \mid x < y\}, \ \Delta := \{(x, y) \in \mathbb{R}^2 \mid x = y\}, \ \langle \mu, f \rangle := \int_{\mathbb{R}^2} f(p) d\mu(p),$$

for any Borel measure μ on the plane, and any $f : \mathbb{R}^2 \to \mathbb{R}$ measurable.

In the following we will use *integrable* measures and *uniformly integrable* sequences of measures. See (Hendrych and Nagy, 2022). Moreover, we make use of weak and vague convergence of measures, which are standard notions in measure theory. See, for instance, (Kallenberg, 1997). We report the definitions in Section A. We write $\mu_n \xrightarrow{w} \mu$ for weak convergence and $\mu_n \xrightarrow{v} \mu$ for vague convergence, Lastly, we also set the following notation: given $p = (x, y) \in \mathbb{R}^2$, we set $(1, p) := (1, x, y) \in \mathbb{R}^3$.

For the sake of simplicity, we use the following simplified definition of persistence diagrams, with all the points having multiplicity 1.

Definition 2 *A persistence diagram is a finite set $D \subset \mathbb{R}^2_{x<y}$. For every persistence diagram D, we can build the positive finite measure $\mu_D = \sum_{p \in D} \delta_p$.*

Following Divol and Lacombe (2021) we give the following definition.

Definition 3 *For any integrable measure μ and for any subset $Z \subset \mathbb{R}^2$, we define:*

$$\mathrm{Pers}_Z(\mu) = \frac{1}{2} \int_Z (y - x) d\mu((x, y)).$$

When $Z = \mathbb{R}^2$, we simply write $\mathrm{Pers}(\mu)$.

As for other functional representation of PDs, see (Adams et al., 2017), we need to re-weight diagrams so that the weight assigned to points goes to zero as we approach Δ.

Definition 4 *A function $\omega : \mathbb{R}^2 \to (0, 1]$ is called a C-stable lift weighting if the function $\Gamma_\omega(p) = \omega(p) \cdot (1, p)$ is C-Lipschitz and satisfies $\| \Gamma_\omega(p) \|_2 \leq C \cdot \frac{y-x}{2}$ for every $p = (x, y) \in \mathbb{R}^2$. Similarly, ω is called an effective (lift) weighting if for any sequence of diagrams $\{\mu_{D_n}\}_{n \in \mathbb{N}}$:*

$$\lim_{r \to \infty} \sup_n \int_{B_r^c} \omega(p) \|p\|_2 \, d\mu_n(p) = 0 \qquad \Longrightarrow \qquad \lim_{r \to \infty} \sup_n \mathrm{Pers}_{B_r^c}(\mu_{D_n}) = 0.$$

Given a persistence diagram D and a weighting ω, we set $\mu_D^\omega := \sum_{p \in D)} \omega(p) \cdot \delta_p$.

Example 5 *Set $\lambda(p) := \frac{y-x}{2\|(1,p)\|_2}$. The following are stable weightings:*

$$\widetilde{\omega}(p) = \lambda(p)^\alpha, \qquad \omega_K^\alpha(p) = \frac{2}{\pi} \arctan\left(\frac{\lambda(p)^\alpha}{K^\alpha}\right),$$

for any $K > 0$ and $\alpha \geq 1$. They are also effective weightings for $\alpha = 1$. See (Pegoraro, 2025) for more details.

As proven in Skraba and Turner (2020), in the context of linear operators defined on spaces of measures, we are forced to work with the 1-Wasserstein metric.

Definition 6 *The 1-Wasserstein distance between persistence diagrams is defined as:*

$$W_1(D, D') = \inf_\gamma \sum_{p \in D_\gamma} \| p - \gamma(p) \|_\infty + \sum_{p \in D - D_\gamma} \| p - \Delta \|_\infty + \sum_{q \in D' - D'_\gamma} \| q - \Delta \|_\infty$$

where $D_\gamma \subset D$, $D'_\gamma \subset D'$, *and* $\gamma : D_\gamma \to D'_\gamma$ *is a bijection. Such a* γ *is usually referred to as a partial matching between* D *and* D'.

Theorem 7 (Divol and Lacombe (2021)) *We have:*

$$W_1(D, D_n) \to 0 \text{ if, and only if, } \mu_{D_n} \xrightarrow{v} \mu_D \text{ and } \mathrm{Pers}(\mu_{D_n}) \to \mathrm{Pers}(\mu_D).$$

2.3. Lift Zonoids of Discrete Measures and Persistence Spheres

We now report the construction of the lift zonoid of a discrete measure, taken from Koshevoy and Mosler (1998); Hendrych and Nagy (2022).

Definition 8 *Given* $\mu = \sum_{i=1}^n c_i \cdot \delta_{p_i}$, $p_i \in \mathbb{R}^2$ *and* $c_i > 0$, *the lift zonoid of* μ *is the following convex set (zonotope):*

$$Z_\mu = \bigoplus_{i=1}^n c_i[0, (1, p_i)] \subset \mathbb{R}^3,$$

with $[0, (1, p_i)]$ *being the segment joining the origin* $0 \in \mathbb{R}^3$ *and the point* $(1, p_i)$.

Koshevoy and Mosler (1998); Hendrych and Nagy (2022) show the following result.

Proposition 9 *We have:*

$$d_H(Z_{\mu_D}, Z_{\mu_{D_n}}) \to 0 \text{ if, and only if, } \mu_{D_n} \xrightarrow{w} \mu_D \text{ and } \{\mu_{D_n}\}_{n\in\mathbb{N}} \text{ is uniformly integrable.}$$

We highlight the following facts:

1. the lift zonoid construction is linear (and injective): $\lambda_1\mu_1 + \lambda_2\mu_2 \mapsto \lambda_1 Z_{\mu_1} \oplus \lambda_2 Z_{\mu_2}$;
2. by linearity, the support function of (the lift zonoid of) a persistence diagram D, with weighting function ω, can be explicitly written as:

$$h_{Z_{\mu_D^\omega}}(v) = \sum_{p\in D} \omega(p)\,\mathrm{ReLu}(\langle v, (1, p)\rangle),$$

where $\mathrm{ReLu}(x) := \max\{0, x\}$.

Definition 10 *Given a persistence diagram* D, *the persistence sphere (PS) of* D *with weighting* ω *is defined as* $\varphi_D^\omega := (h_{Z_{\mu_D^\omega}})_{|\mathbb{S}^2}$.

Remark 11 *The domain of a persistence sphere is the unit sphere* $\mathbb{S}^2$. *Indeed, a persistence diagram is a subset of* $\mathbb{R}^2$, *hence its lift zonoid lives in* $\mathbb{R}^3$. *Therefore, its support function is completely determined by its restriction to* $\mathbb{S}^2$.

3. CONTINUITY THEOREMS

We now state our main results. Theproofs can be found in Pegoraro (2025).

Theorem 12 *Let μ_D, μ'_D be PDs and let $\omega : \mathbb{R}^2 \to \mathbb{R}$ be a stable weighting. We have:*

$$d_H(Z_{\mu_D^\omega}, Z_{\mu_{D'}^\omega}) \leq \max\{C, C'\} \cdot W_1(\mu_D, \mu_{D'}),$$

with $C, C' > 0$ being the stability constants of ω (see Theorem 4).

Theorem 13 *Let $\{\mu_{D_n}\}_{n\in\mathbb{N}}$ be a sequence of PDs such that $d_H(Z_{\mu_{D_n}^\omega}, Z_{\mu_D^\omega}) \to 0$, with $\omega : \mathbb{R}^2 \to \mathbb{R}$ being an effective weighting. Then, $W_1(\mu_{D_n}, \mu_D) \to 0$.*

Summarizing the statements of Theorem 12 and Theorem 13, and writing them replacing lift zonoids with persistence spheres (using Theorem 1), and using the bound $\int_{\mathbb{S}^2} |f|^p d\mathcal{L} \leq \mathcal{L}(\mathbb{S}^2) \sup_{v\in\mathbb{S}^2} |f(v)|^p$ we obtain the following result.

Corollary 14 *Within the setting of the previous results, we have:*

- *for any $p \in [1, \infty]$ there exist $C_p > 0$ such that, for every pair of diagrams D, D', we have $\| \varphi_D^\omega - \varphi_{D'}^\omega \|_p \leq C_p \cdot W_1(D, D')$;*
- *if $\| \varphi_D^\omega - \varphi_{D_n}^\omega \|_\infty \to 0$, then $W_1(D_n, D) \to 0$.*

4. EXPERIMENTS

We tested PSs on several (regression and classification) case studies. We compared their performances against persistence images (PIs), persistence landscapes (PLs), and the sliced Wasserstein Kernel (SWK). For PSs, PIs, and PLs we considered both support vector machine (SVM) methods with radial-basis kernels ("-SMV" in Table 1), and (penalized) linear and logistic regressions ("-Lin" in Table 1). Performances were measured in terms of average R^2 for regression and average accuracy for classification. See also (Pegoraro, 2025) for additional experimental results.

5. CONCLUSION

In this extended abstract, we introduced a novel functional representation of persistence diagrams and outlined preliminary results that highlight its advantages over existing summaries. These theoretical findings are supported by experiments that demonstrate the effectiveness of our approach. However, further investigation is needed into aspects such as the interpretability of the proposed summary and the impact of different weighting functions.

Table 1: Results of the case studies: we report average R^2 for regression and average accuracy for classification. Between brackets we reported the standard deviation of the scores across the 10 repetitions. Best performing pipelines are reported in bold.

	PS-SVM	PS-Lin	PI-SVM	PI-Lin	PL-SVM	PL-Lin	SWK
Regression							
Eyeglasses Case Study	0.970	**0.978**	0.684	0.487	0.956	0.878	0.971
Tecator	**0.962 (0.004)**	0.960 (0.007)	0.928 (0.029)	0.922 (0.018)	0.939 (0.016)	0.837 (0.048)	0.953 (0.010)
Classification							
Growth	**0.864 (0.052)**	0.860 (0.067)	0.525 (0.095)	0.536 (0.086)	0.782 (0.051)	0.767 (0.107)	0.768 (0.058)
NOx	**0.846 (0.068)**	**0.846 (0.048)**	0.803 (0.073)	0.563 (0.102)	0.794 (0.047)	0.751 (0.040)	0.840 (0.055)
DYN_SYS	0.799 (0.040)	0.823 (0.027)	0.709 (0.020)	0.819 (0.017)	**0.849 (0.029)**	0.823 (0.027)	0.828 (0.028)
ENZYMES_JACC	**0.292 (0.048)**	0.276 (0.023)	-	-	0.236 (0.026)	0.254 (0.028)	0.283 (0.055)
POWER	0.761 (0.021)	0.733 (0.019)	0.682 (0.026)	0.700 (0.020)	0.746 (0.014)	0.719 (0.030)	**0.767 (0.15)**
SHREC14	0.818 (0.046)	0.910 (0.024)	0.905 (0.029)	0.897 (0.024)	0.914 (0.021)	**0.931 (0.013)**	0.886 (0.092)
Human Poses	0.575 (0.081)	**0.640 (0.094)**	0.407 (0.144)	0.503 (0.046)	0.460 (0.118)	0.480 (0.105)	0.345 (0.082)
McGill 3D Shapes	0.411 (0.086)	0.489 (0.173)	**0.678 (0.060)**	0.045 (0.052)	0.650 (0.093)	0.511 (0.150)	0.567 (0.13)

References

Henry Adams, Tegan Emerson, Michael Kirby, Rachel Neville, Chris Peterson, Patrick Shipman, Sofya Chepushtanova, Eric Hanson, Francis Motta, and Lori Ziegelmeier. Persistence images: A stable vector representation of persistent homology. *Journal of Machine Learning Research*, 18(8):1–35, 2017.

Dashti Ali, Aras Asaad, Maria-Jose Jimenez, Vidit Nanda, Eduardo Paluzo-Hidalgo, and Manuel Soriano-Trigueros. A survey of vectorization methods in topological data analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(12):14069–14080, 2023.

Aras Asaad, Dashti Ali, Taban Majeed, and Rasber Rashid. Persistent homology for breast tumor classification using mammogram scans. *Mathematics*, 10(21):4039, 2022.

Cinzia Bandiziol and Stefano De Marchi. Persistence symmetric kernels for classification: A comparative study. *Symmetry*, 16(9):1236, 2024.

Christophe AN Biscio and Jesper Møller. The accumulated persistence function, a new useful functional summary statistic for topological data analysis, with a view to brain artery trees and spatial point process applications. *Journal of Computational and Graphical Statistics*, 28(3):671–681, 2019.

Michael M Bronstein and Iasonas Kokkinos. Scale-invariant heat kernel signatures for non-rigid shape recognition. In *2010 IEEE computer society conference on computer vision and pattern recognition*, pages 1704–1711. IEEE, 2010.

Peter Bubenik. Statistical topological data analysis using persistence landscapes. *Journal of Machine Learning Research*, 16:77–102, 2015.

Mathieu Carriere, Marco Cuturi, and Steve Oudot. Sliced wasserstein kernel for persistence diagrams. In *International conference on machine learning*, pages 664–673. PMLR, 2017.

Barbara Di Fabio and Massimo Ferri. Comparing persistence diagrams through complex vectors. In *International conference on image analysis and processing*, pages 294–305. Springer, 2015.

Vincent Divol and Théo Lacombe. Understanding the topology and the geometry of the space of persistence diagrams via optimal partial transport. *Journal of Applied and Computational Topology*, 5:1–53, 2021.

Herbert Edelsbrunner and John L Harer. *Computational topology: an introduction*. American Mathematical Society, 2010.

Manuel Febrero, Pedro Galeano, and Wenceslao González-Manteiga. Outlier detection in functional data by depth measures, with application to identify abnormal nox levels. *Environmetrics*, 19(4):331–345, 2008.

Frédéric Ferraty and Philippe Vieu. *Nonparametric functional data analysis: theory and practice*. Springer Verlag, NY, 2006. URL https://doi.org/10.1007/0-387-36620-2.

Vesna Gotovac Dogaš and Marcela Mandarić. Topological data analysis for random sets and its application in detecting outliers and goodness of fit testing. *Statistical Methods & Applications*, pages 1–45, 2025.

František Hendrych and Stanislav Nagy. A note on the convergence of lift zonoids of measures. *Stat*, 11(1):e453, 2022.

Sara Kališnik. Tropical coordinates on the space of persistence barcodes. *Foundations of Computational Mathematics*, 19(1):101–129, 2019.

Olav Kallenberg. *Foundations of modern probability*. Springer, 1997.

Gleb Koshevoy and Karl Mosler. Lift zonoids, random convex hulls and the variability of random vectors. 1998.

Genki Kusano, Kenji Fukumizu, and Yasuaki Hiraoka. Kernel method for persistence diagrams via kernel embedding and weight factor. *Journal of Machine Learning Research*, 18(189):1–41, 2018.

Yuriy Mileyko, Sayan Mukherjee, and John Harer. Probability measures on the space of persistence diagrams. *Inverse Problems*, 27(12):124007, 2011.

Anthea Monod, Sara Kalisnik, Juan Ángel Patino-Galindo, and Lorin Crawford. Tropical sufficient statistics for persistent homology. *SIAM Journal on Applied Algebra and Geometry*, 3(2):337–371, 2019.

Steve Y Oudot. *Persistence theory: from quiver representations to data analysis*, volume 209. American Mathematical Society Providence, 2015.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Matteo Pegoraro. Persistence spheres: Bi-continuous representations of persistence diagrams. *arXiv preprint arXiv:2509.16999*, 2025.

Matteo Pegoraro and Mario Beraha. Projected statistical methods for distributional data on the real line with the wasserstein metric. *Journal of Machine Learning Research*, 23 (37):1–59, 2022.

The GUDHI Project. *GUDHI User and Reference Manual*. GUDHI Editorial Board, 3.11.0 edition, 2025. URL https://gudhi.inria.fr/doc/3.11.0/.

Chi Seng Pun, Si Xian Lee, and Kelin Xia. Persistent-homology-based machine learning: a survey and a comparative study. *Artificial Intelligence Review*, 55(7):5169–5213, 2022.

Carlos Ramos-Carreño, José L. Torrecilla, Miguel Carbajo Berrocal, Pablo Marcos Manchón, and Alberto Suárez. scikit-fda: A Python Package for Functional Data Analysis. *Journal of Statistical Software*, 109(2):1–37, May 2024.

Jan Reininghaus, Stefan Huber, Ulrich Bauer, and Roland Kwitt. A stable multi-scale kernel for topological machine learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4741–4748, 2015.

R Tyrrell Rockafellar. *Convex Analysis*, volume 28. Princeton University Press, 1997.

Gabriella Salinetti and Roger J.-B. Wets. On the convergence of sequences of convex sets in finite dimensions. *SIAM Review*, 21(1):18–33, 1979.

Nathaniel Saul and Chris Tralie. Scikit-tda: Topological data analysis for python, 2019. URL https://doi.org/10.5281/zenodo.2533369.

Primoz Skraba and Katharine Turner. Wasserstein stability for persistence diagrams. *arXiv preprint arXiv:2006.16824*, 2020.

RD Tuddenham and MM Snyder. Physical growth of california boys and girlsfrom birth to age 18. *Calif. Publ. Child Develop*, 1:183–364, 1954.

Valeria Vitelli, Laura Maria Sangalli, Piercesare Secchi, and Simone Vantini. Functional clustering and alignment methods with applications. *Communications in Applied and Industrial Mathematics*, 1(1):205–224, 2010. URL https://doi.org/10.1685/2010CAIM486.

Appendix A. Additional Preliminaries

In this section we report some well-known definitions, taken from the references provided in the main text, which can help the reader in following the extended abstract.

A.1. Convex Sets and Support Functions

Definition A.1 *Given to convex sets $A, B \subset \mathbb{R}^2$, their Minkowski sum and their multiplication with a non-negative scalar $\lambda \geq 0$, are defined as:*

$$A \oplus B = \{a + b \mid a \in A, b \in B\}, \lambda A = \{\lambda a \mid a \in A\}.$$

Definition A.2 *Given a compact convex set $A \subset \mathbb{R}^2$, its support function is defined as:*

$$h_A : \mathbb{R}^2 \to \mathbb{R}$$
$$x \mapsto \max_{a \in A} \langle x, a \rangle.$$

Definition A.3 *Given two compact subsets $A, B \subset Z$, with (Z, d_Z) being a metric space, their Hausdorff distance is defined as:*

$$d_H(A, B) = \max\{\max_{a \in A} d_Z(a, B), \max_{b \in B} d_Z(b, A)\}$$

A.2. Persistence Diagrams and Measures

Definition A.4 *A positive finite Borel measure on $\mathbb{R}^2$, μ, is called integrable if:*

$$\langle \mu, \| \cdot \|_2 \rangle = \int_{\mathbb{R}^2} \| p \|_2 \, d\mu(p) < \infty.$$

Similarly, a sequence of integrable measures $\{\mu_n\}_{n \in \mathbb{N}}$ is uniformly integrable if:

$$\lim_{r \to \infty} \sup_{n \in \mathbb{N}} \int_{\|p\|_2 > r} \| p \|_2 \, d\mu_n(p) = 0.$$

Definition A.5 *A sequence of integrable measures $\{\mu_n\}_{n \in \mathbb{N}}$ converges weakly to μ if:*

$$\langle \mu_n, f \rangle \to \langle \mu, f \rangle,$$

for every $f : \mathbb{R}^2 \to \mathbb{R}$ continuous and bounded. If $\langle \mu_n, f \rangle \to \langle \mu, f \rangle$ for every $f : \mathbb{R}^2 \to \mathbb{R}$ continuous and compactly supported, we say that $\{\mu_n\}_{n \in \mathbb{N}}$ converges vaguely to μ.

Appendix B. Case Studies Details

We now give a brief description of the considered case studies.

The "Eyeglasses" dataset is a regression case study we designed using the *eyeglasses* generative model from the `scikit-tda` python package (Saul and Tralie, 2019). This model takes two radii as parameters, and a noise variable which was kept equal to 1. The first radius was always set equal to 20, while the second was sampled according to a normal distribution with mean 10 and standard variation 2.5. We sampled 2000 point clouds and used a $30\% - 70\%$ split between training and test data; threefold cross-validation was used to select hyper-parameters.

For the following functional datasets we used a $70\% - 30\%$ split between training and test data and threefold cross-validation was used to select hyper-parameters. All the datasets are freely available within the `scikit-fda` python package (Ramos-Carreño et al., 2024).

The "Tecator" dataset (https://lib.stat.cmu.edu/datasets/tecator) consists of publicly available measurements collected using the "Tecator Infratec Food and Feed Analyzer". Building on the derivatives of these curves, we explore the same regression problem as in Ferraty and Vieu (2006), trying to regress the fat content of the food samples.

The "NO_x" dataset (Febrero et al., 2008) contains hourly measurements of daily nitrogen oxides (NO_x) emissions in the Barcelona area. The data is labeled based on whether the emission curve was recorded on a weekday or a weekend, and our goal is thus to reconstruct this labeling through supervised classification.

The "Growth" dataset (Tuddenham and Snyder, 1954), also known as "The Berkeley Growth Study", contains height measurements of girls and boys, recorded yearly between ages 1 and 18. A common approach is to analyze the first derivative of the growth curves to distinguish growth dynamics between boys and girls (Vitelli et al., 2010).

The datasets, "DYN SYS", "ENZYMES JACC", "POWER", and "SHREC14", instead, were taken from Bandiziol and De Marchi (2024): see (Bandiziol and De Marchi, 2024) for a description of the classification problems. The persistence diagrams associated to the classification problems therein described are available at: https://github.com/cinziabandiziol/persistence_kernels.

Lastly, "Human Poses" and "Mc Gill 3D Shapes", where taken from the webpage https://github.com/ctralie/TDALabs/blob/master/3DShapes.ipynb. The classification pipelines are detailed in the referenced notebook and consist of the following: a sublevel set filtration of the height function for the human pose classification task, and a sublevel set filtration of the heat kernel signature (Bronstein and Kokkinos, 2010) for the 3D shape classification case study. For both datasets we used a $80\% - 20\%$ split between training and test data and threefold cross-validation was used to select hyper-parameters. We note that the "Mc Gill 3D Shapes" dataset is actually a subsample from the original "Mc Gill 3D Shapes", dataset used in our experiments is a subsample of the original dataset, which we were no longer able to access in its complete form online. In both case studies, the train-test split was determined by the dataset limitation of having only 10 samples per class.

Appendix C. Implementation Details

As mentioned in the main text, we considered two types of pipelines based on the vectorization methods: support vector machines (SVM) and penalized linear/logistic regression. Both model types were implemented using the `scikit-learn` Python package (Pedregosa et al., 2011). In particular, for linear and logistic regression we considered the ℓ_2 penalization to avoid overfitting.

We first list the different hyperparameters we had to set for each pipeline ("-SVM" vs "-Lin")and then the hyperparameters which are bound to each different linearization method.

- -SVM: the SVM methods all needed the choice of a regularization hyperparameter C. It was chosen among the following values $\{0.001, 0.01, 0.1, 1, 10, 100, 1000, 10000\}$. Apart from the SWK method, for which the kernel was precomputed, for all other methods, the radial basis function kernel was used;
- -Lin: the penalization parameter α for ridge regression was chosen within $\{0.01, 0.1, 1, 10, 100\}$; while the C parameter for penalized logistic regression was chosen among $\{1, 10, 100, 1000, 10000\}$.

The different linearization methods required the following parameters:

- PS: we used the ω_K^α weighting function we define in the main text, and we selected $\alpha \in \{1, 2, 4, 8, 12, 16\}$ and $K \in \{0.0001, 0.001, 0.01, 0.1, 0.2, 0.3, 0.4\}$. Since PS are functions defined on $\mathbb{S}^2$but rather on spherical coordinates, we projected the data onto a spline basis and applied functional principal component analysis (FPCA) on the sphere—retaining all components—to obtain a vectorized representation in an orthonormal basis suitable for penalized methods from `scikit-learn`. FPCA was implemented using a Rayleigh quotient approach, which works well for non-orthonormal bases. For a related example, see (Pegoraro and Beraha, 2022).
- PI: we used the `scikit-tda` module `persim`, searching its hyperparameters in the following sets: the *pixel_size* was selected by taking the rectangle (in the birth-persistence coordinates) containing all the PDs, and then dividing its shortest side by n', $n' \in \{100, 500\}$. Some suitable rounding to the closest power of 10 was used to avoid numerical inconsistencies. We used the default bivariate normal Gaussian kernel, with σ being a fraction of the *pixel_size*, obtained as $pixel\_size/m$, $m \in \{1, 10, 100\}$. Lastly, the parameter n in the *weight_params* dictionary (for the weight given by *persistence*) was considered in $\{1, 2, 4, 8\}$.
- PL: for persistence landscapes, we considered the first 5 landscapes, we evaluated all of them on a common grid (fixed for all the elements in the dataset) and we concatenated them. Thus, we did not have to set any hyperparameters.
- SWK: we used the implementation of the sliced Wasserstein distance in the `gudhi` library (Project, 2025). The parameter M was kept equal to 100, while we selected $\sigma \in \{0.00001, 0.0001, 0.001, 0.01, 0.1, 1, 10\}$ for the Graham matrix.