Differentiability and Optimization of Multiparameter Persistent Homology

Luis Scoccola [*1] **Siddharth Setlur** [*2] **David Loiseaux** [3] **Mathieu Carrière** [3] **Steve Oudot** [4]

Abstract

Real-valued functions on geometric data—such as node attributes on a graph—can be optimized using descriptors from persistent homology, allowing the user to incorporate topological terms in the loss function. When optimizing a single real-valued function (the one-parameter setting), there is a canonical choice of descriptor for persistent homology: the barcode. The operation mapping a real-valued function to its barcode is differentiable almost everywhere, and the convergence of gradient descent for losses using barcodes is relatively well understood. When optimizing a vector-valued function (the multiparameter setting), there is no unique choice of descriptor for multiparameter persistent homology, and many distinct descriptors have been proposed. This calls for the development of a general framework for differentiability and optimization that applies to a wide range of multiparameter homological descriptors. In this article, we develop such a framework and show that it encompasses well-known descriptors of different flavors, such as signed barcodes and the multiparameter persistence landscape. We complement the theory with numerical experiments supporting the idea that optimizing multiparameter homological descriptors can lead to improved performances compared to optimizing one-parameter descriptors, even when using the simplest and most efficiently computable multiparameter descriptors.

[*]Equal contribution [1]Mathematical Institute, University of Oxford, UK [2]Department of Mathematics, ETH Zürich, Switzerland [3]DataShape, Centre Inria d'Université Côte d'Azur, France [4]GeomeriX, Inria Saclay and École polytechnique, Paris, France. Correspondence to: Luis Scoccola <luis.scoccola@maths.ox.ac.uk>, Siddharth Setlur <ssetlur@student.ethz.ch>.

Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024. Copyright 2024 by the author(s).

1 Introduction

Context. Persistent homology (PH), the main tool of topological data analysis (TDA), produces a descriptor in the form of a *barcode* for real-valued *filtering functions*, i.e., real-valued functions on graphs (more generally, simplicial complexes) that assign higher values to edges than to their vertices (more generally, monotonic with respect to the face order). PH has found a variety of applications in machine learning (ML), notably: as a featurization technique encoding complementary information to traditional descriptors (see, e.g., (Hensel et al., 2021) for a survey), offering stability guarantees with respect to perturbations of the data (Cohen-Steiner et al., 2007), and the possibility for end-to-end feature learning (Hofer et al., 2020b; Carrière et al., 2021); as a topological regularizer for constraining models to follow some prescribed topology in order to avoid, e.g., overfitting (Chen et al., 2019; Moor et al., 2020); and as a topological layer in neural networks for enhancing network performance (Carrière et al., 2020; Zhao et al., 2020a; Kim et al., 2020). A sound mathematical foundation for this type of applications, and for barcode optimization in general, was developed in (Carrière et al., 2021).

Recent contributions in the area of TDA for ML have demonstrated empirically the added value, in terms of learning performances, of replacing real-valued filtering functions by $\mathbb{R}^n$-valued filtering functions in the pipelines involving PH; see, e.g., (Carrière & Blumberg, 2020; Demir et al., 2022; Loiseaux et al., 2023b; Chen et al., 2023). However, in contrast to the usual real-valued setting, there does not exist—and, due to fundamental algebraic reasons, there *cannot* exist (Carlsson & Zomorodian, 2009; Bauer & Scoccola, 2023)—any discrete descriptor like the barcode that completely encodes the PH of $\mathbb{R}^n$-valued filtering functions. As a consequence, the aforementioned contributions resorted to a variety of incomplete descriptors, such as *multiparameter persistence landscapes* (Vipond, 2020) and generalizations (Xin et al., 2023), *signed barcodes as measures* (Loiseaux et al., 2023b), *multiparameter persistent images* (Carrière & Blumberg, 2020; Loiseaux et al., 2023a), or *multiparameter persistence kernels* (Corbet et al., 2019).

In order to generalize the use of these incomplete descriptors to other applications in machine learning, it is necessary to extend the existing optimization framework to them. To

the best of our knowledge, this has not been done yet, but we expect to see it happen in the near future. However, the diversity of the proposed descriptors is likely to induce the parallel development of multiple competing, specialized frameworks, depriving the field of a general, unified theory.

In order to prevent this undesirable situation we aim at developing optimization frameworks for classes of descriptors instead of single descriptors. The problem is then to find the right level of abstraction to prove a differentiability and convergence result, that is, a general enough – yet still easily checkable – set of assumptions, under which these seemingly very different invariants can be proven to be differentiable with gradient descent convergence guarantees.

The present paper is a first attempt at this, focusing on the class of descriptors characterized as being *semilinearly determined on grids*. A formal statement of this condition is given in Section 3 (Definition 3.17), but intuitively, it means that the descriptor, viewed as a map from the space $\mathsf{Fil}_n(K)$ (of $\mathbb{R}^n$-valued filtering functions on a fixed simplicial complex K) to a Euclidean space $\mathbb{R}^D$ factors through a simple linear space consisting of inclusions of grids in $\mathbb{R}^n$, where it behaves semilinearly (in the sense of piecewise affinely). This property is directly inspired from the well-established fact that the PH of real-valued filtering functions itself can be computed over finite integer grids, and behaves affinely with respect to grid inclusions: indeed, the barcodes produced by the persistence algorithm (Edelsbrunner et al., 2002; Zomorodian & Carlsson, 2005) depend affinely on the input function values as long as the pairing of the simplices in K remains fixed, and the pairing itself is entirely prescribed by the pre-order on the simplices induced by the function values. While we cannot assert that all future descriptors for $\mathbb{R}^n$-valued filtering functions will satisfy our condition, we show that currently known descriptors of very different flavors, such as the signed barcodes as measures and the multiparameter persistence landscapes, do.

Theoretical contributions. Theorem 1.1, below, summarizes our contributions regarding general homological descriptors of multiparameter filtrations (Theorem 3.19 and Proposition 3.20). It generalizes the existing optimization framework for barcodes of real-valued filtering functions (Carrière et al., 2021) to descriptors of $\mathbb{R}^n$-valued filtering functions that are semilinearly determined on grids.

Theorem 1.1. *Assume given a simplicial complex K, a number of parameters $n \in \mathbb{N}_{\geq 1}$, and a descriptor $\alpha\colon \mathsf{Fil}_n(K) \to \mathbb{R}^D$. If α is semilinearly determined on grids, then it is a semilinear map, with explicit Clarke subdifferential. If furthermore α is included in an optimization pipeline of the following form, where Φ is a parametrized family of fitrations and where E is some loss function:*

$$\mathbb{R}^d \xrightarrow{\Phi} \mathsf{Fil}_n(K) \xrightarrow{\alpha} \mathbb{R}^D \xrightarrow{E} \mathbb{R},$$

such that the composite objective function $\mathcal{L} := E \circ \alpha \circ \Phi$ is locally Lipschitz, and the maps Φ and E are definable over a common o-minimal structure, then, under the usual assumptions of the stochastic subgradient method (Assumption B.1), almost surely, every limit point of the iterates of the stochastic subgradient method is critical for the objective function $\mathcal{L}$, and the sequence of values of $\mathcal{L}$ converges.

For illustrative purposes, Theorem 1.1 is stated for a pipeline using a single (topology-based) loss. Several generalizations are possible and we comment on these in Remark B.3.

To illustrate the flexibility of our approach, we apply our framework to two widely different topology-based descriptors: the signed barcodes as measures (Loiseaux et al., 2023b) derived from the Hilbert function (Oudot & Scoccola, 2024) and the multiparameter persistence landscape (Vipond, 2020). As in the one-parameter case, these descriptors take values in infinite dimensional spaces, and thus we prove semilinearity for suitable finite dimensional representations of them (see Appendix D for details).

Theorem 1.2. *The sorted Hilbert decomposition and the evaluated multiparameter persistence landscape are semilinearly determined on grids.*

The locally Lipschitzness condition of Theorem 1.1 can then be satisfied using Proposition E.4, in the case of the Hilbert decomposition signed measure, and the stability theorem in (Vipond, 2020), in the case of multiparameter persistence landscapes.

Analogous results can be obtained for other invariants, such as for instance signed measures derived from the rank invariant (Definition A.3).

See Figure 8 for a dependency graph of the main definitions, results, and experiments in the paper, and Table 2 for a notation table.

Practical contributions. Examples 3.22 and 3.23, in Section 3.3, give concrete instantiations of our theory suited for end-to-end optimization using signed barcodes as measures. In Section 4 we report on numerical experiments showcasing these and other pipelines, on both synthetic and real data. The main conclusion is that, for machine learning purposes, multiparameter homological descriptors tend to outperform their one-parameter counterpart.

Related work. As already mentioned, and to the best of our knowledge, the general problem of establishing the differentiability and gradient descent convergence of multiparameter persistence has not been addressed in the literature so far. Concurrent work (Mukherjee et al., 2024) addresses this problem in the special case of GRIL (Xin et al., 2023), a generalization of the multiparameter persistence landscape, and develops a specific, yet elementary, optimization framework for GRIL.

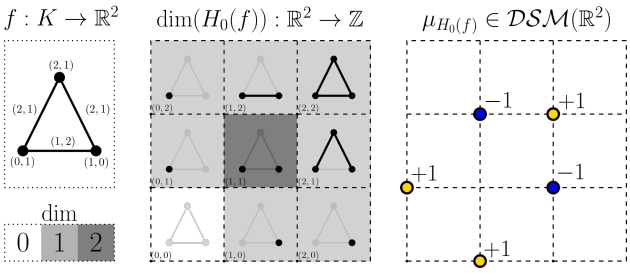


Figure 1. An example, from (Loiseaux et al., 2023b), of a filtered simplicial complex, the Hilbert function of its 0th persistent homology, and the corresponding Hilbert decomposition signed measure.

The one-parameter case is well studied (Carrière et al., 2021; Leygonie et al., 2023; 2022), with applications including shape matching and classification (Hofer et al., 2017; Poulenard et al., 2018), representation learning (Hofer et al., 2019), graph classification (Hofer et al., 2020b; Horn et al., 2022a; Zhao et al., 2020b), and regularization (Chen et al., 2019; Hofer et al., 2020a); see also Sections 3.2.3 and 3.3.1 of (Hensel et al., 2021) for an overview, but note that, at this point, any list will necessarily be incomplete.

2 Background

This section can be skipped and only referred to as required, depending on the background of the reader. For clarity and conciseness, we do not give full details of some constructions that are not used in the main results of this paper (see Appendix A for more background). Whenever we skip details we warn the reader and provide a precise reference.

Basic notation. If $n \geq 1 \in \mathbb{N}$ and X is a finite set, we let $(\mathbb{R}^n)^X$ denote the set of functions $X \to \mathbb{R}^n$, which is a real vector space of dimension $n \times |X|$.

Simplicial complexes. A finite *simplicial complex* K consists of a finite set X and a collection $K \subseteq \mathsf{parts}(X) \setminus \{\emptyset\}$ of non-empty subsets X with the property that $\{x\} \in K$ for every $x \in X$, and if $\sigma \in K$ and $\tau \subseteq \sigma$, then $\tau \in K$. Let $i \in \mathbb{N}$; the i-simplices of K are the sets $\sigma \in K$ such that $|\sigma| = i + 1$. In particular, the 0-simplices correspond exactly to the elements of the underlying set X. For this reason, it is common to denote a simplicial complex by K, leaving the underlying set X implicit.

Note that any simplicial complex K is automatically a partially ordered set where $\tau \leq \sigma \in K$ precisely when $\tau \subseteq \sigma$. This order is known as the *face order* of K.

Example 2.1. Any finite, simple, undirected graph G is equivalently a simplicial complex with 0-simplices the vertices of G, 1-simplices the edges of G, and no higher-dimensional simplices.

Filtrations. Let K be a finite simplicial complex.

Definition 2.2. Let $n \in \mathbb{N}$. An *n-filtration* on K consists of a function $f : K \to \mathbb{R}^n$, which is monotonic with respect to the face order of K and product (partial) order on $\mathbb{R}^n$.

Unraveling definitions, an n-filtration is a function $f : K \to \mathbb{R}^n$ mapping simplices of K to vectors in $\mathbb{R}^n$, with the property that for each pair of simplices $\sigma \subseteq \tau \in K$ and $1 \leq i \leq n$, we have $f_i(\sigma) \leq f_i(\tau)$, where $f_i : K \to \mathbb{R}$ denotes the ith coordinate of f. See Figure 1 for an illustration.

We let $\mathsf{Fil}_n(K)$ denote the set of n-filtrations of K. Note that $\mathsf{Fil}_n(K) \subseteq (\mathbb{R}^n)^K$.

Example 2.3. Let G be an undirected graph and let $f' : \mathsf{vert}(G) \to \mathbb{R}^n$ be any function. One can extend f' to a filtering function $f \in \mathsf{Fil}_n(G)$ by mapping a 0-simplex $\{x\}$ (corresponding to a vertex x) to $f'(x) \in \mathbb{R}^n$, and a 1-simplex $\{x, y\}$ (corresponding to an edge between x and y) to $(\max(f'(x)_1, f'(y)_1), \ldots, \max(f'(x)_n, f'(y)_n)) \in \mathbb{R}^n$. This is an instance of a lower-star filtration (Edelsbrunner & Harer, 2008), a common construction in TDA.

(Persistent) homology. For detailed introductions to one- and multiparameter persistence, see, e.g., (Edelsbrunner & Harer, 2008) and (Botnan & Lesnick, 2023). Fix a field $\mathbb{F}$ for the rest of this paper, let $i \in \mathbb{N}$, and let K be a simplicial complex.

The ith *homology* of K with coefficients in $\mathbb{F}$ is an $\mathbb{F}$-vector space $H_i(K)$; see, e.g., (Fomenko, 1994) for a precise definition with illustrations. Informally, the dimension of $H_i(K)$ counts the number of independent i-dimensional holes of K; as an example, the dimension of $H_0(K)$ equals the number of connected components of K.

The homology construction is functorial, which in particular implies that if $K \subseteq K'$ are simplicial complexes, there is an $\mathbb{F}$-linear map $H_i(K) \to H_i(K')$ of $\mathbb{F}$-vector spaces. Then, given $f \in \mathsf{Fil}_n(K)$, we can filter K as follows: for $r \in \mathbb{R}^n$, let $K_r^f := \{\sigma \in K : f(\sigma) \leq r\} \subseteq K$. Since f is a filtration, it's clear that $K_r^f \subseteq K_s^f$ whenever $r \leq s \in \mathbb{R}^n$, where, again, $\leq$ denotes the product (partial) order on $\mathbb{R}^n$. This allows for the following construction. For convenience of notations later on, we let $H_i(f)(r) := H_i(K_r^f)$.

Definition 2.4. Let $f \in \mathsf{Fil}_n(K)$ and let $i \in \mathbb{N}$. The ith *multiparameter persistent homology* of f, denoted $H_i(f)$, consists of the collection of vector spaces $\{H_i(f)(r)\}_{r \in \mathbb{R}^n}$ and the linear maps $\{\varphi_{r,s}^f : H_i(f)(r) \to H_i(f)(s)\}_{r \leq s \in \mathbb{R}^n}$, known as the *structure maps*.

See Figure 1 for an illustration.

Let $f \in \mathsf{Fil}_n(K)$. The ith *Hilbert function* of f is the function $\mathsf{Hil}(H_i(f)) : \mathbb{R}^n \to \mathbb{Z}$ defined as the pointwise dimension of $H_i(f)$, that is

$$\mathsf{Hil}(H_i(f))(r) := \dim(H_i(f)(r)).$$

See Figure 1 for an illustration.

The ith *rank invariant* of f is the function $\mathsf{rk}(H_i(f)) : \{(r,s) \in (\mathbb{R}^n)^2 : r \leq s\} \to \mathbb{Z}$ defined as the rank of the structure maps of $H_i(f_i)$, that is

$$\mathsf{rk}(H_i(f))(r,s) := \mathsf{rk}(\varphi^f_{r,s}).$$

The rank invariant has no less information than the Hilbert function: $\mathsf{rk}(H_i(f))(r,r) = \mathsf{Hil}(H_i(f))(r)$ for $r \in \mathbb{R}^n$.

A key motivation for using homology-based descriptors is that they are automatically isomorphism-invariant (Appendix A.3), meaning that they do not depend on, e.g., arbitrary labelings of data.

Discrete signed measures. We define discrete signed measures, which are needed for the definition of the Hilbert decomposition signed measure (Loiseaux et al., 2023b).

If M is a metric space, the set of *discrete measures* on M, denoted $\mathcal{DM}(M)$, is the set of all finite, positive, integer linear combination of Dirac masses on M. The set of discrete signed measures on M, denoted $\mathcal{DSM}(M)$, consists of the set of finite, non-necessarily positive, integer linear combinations of Dirac masses on M. As usual, one can endow the set of discrete signed measures on M the optimal transport distance (Appendix A.1.1), denoted OT, which is an extended metric. If $\mu \in \mathcal{DSM}(M)$, we let μ^+ and μ^- denote its positive and negative part of μ, respectively, according to its Jordan decomposition (see, e.g., p. 421 of (Billingsley, 1995)).

Homological descriptors of multiparameter filtrations. A *descriptor* of filtrations of K consists of a set A and a function $\mathsf{Fil}_n(K) \to A$.

We now define the Hilbert decomposition signed measure (originally introduced in (Loiseaux et al., 2023b)), which is a descriptor of multiparameter filtrations. This descriptor completely characterizes the Hilbert function of a filtration as a discrete signed measure on $\mathbb{R}^n$; this is remarkable, since the Hilbert function is an element of the space of functions $\mathbb{R}^n \to \mathbb{R}$, an infinite dimensional space. For the definition of the rank decomposition signed measure, a stronger descriptor that strictly generalizes the one-parameter barcode, and that completely characterizes the rank invariant of filtrations, see Appendix A.2.

Definition 2.5. Let $f \in \mathsf{Fil}_n(K)$ and let $i \in \mathbb{N}$. The ith *Hilbert decomposition signed measure* is the unique measure $\mu^{\mathsf{Hil}}_{H_i(f)} \in \mathcal{DSM}(\mathbb{R}^n)$ such that, for all $r \in \mathbb{R}^n$,

$$\dim(H_i(f)(r)) = \mu^{\mathsf{Hil}}_{H_i(f)}\left(\{s \in \mathbb{R}^n : s \leq r\}\right).$$

In order to make $\mathbb{R}^n$ into a metric space, so that the optimal transport distance on $\mathcal{DSM}(\mathbb{R}^n)$ is defined, we use $\| - \|_\infty$.

See Figure 1 for an illustration.

Semilinear functions and o-minimal structures. We now introduce semilinear and semialgebraic functions. A generalization of these concepts is that of functions that are defined over an o-minimal structure. For details about this theory, we refer the reader to (van den Dries, 1998). We care about these types of well-behaved functions because gradient descent can be applied on them with convergence guarantees (Davis et al., 2020).

A set $S \subseteq \mathbb{R}^n$ is *semilinear* (resp. *semialgebraic*) if it is a finite union of sets of the form $\{(x_1,\dots,x_n) \in \mathbb{R}^n : P(x_1,\dots,x_n) = 0\}$ with P a polynomial degree ≤ 1 (resp. of any degree), and sets of the form $\{(x_1,\dots,x_n) \in \mathbb{R}^n : P(x_1,\dots,x_n) > 0\}$, with P a polynomial degree ≤ 1 (resp. of any degree). If $S \subseteq \mathbb{R}^n$ is semilinear (resp. semialgebraic), a function $\psi : S \to \mathbb{R}^m$ is *semilinear* (resp. *semialgebraic*) if its graph $\{(x, \psi(x)) : x \in S\} \subseteq \mathbb{R}^n \times \mathbb{R}^m$ is a semilinear (resp. semialgebraic) set. Note that the definitions above also make sense when $\mathbb{R}^n$ and $\mathbb{R}^m$ are replaced with any real vector spaces of finite dimension, since any such vector space can be identified with $\mathbb{R}^n$ using an arbitrary linear isomorphism, and any choice of linear isomorphism leads to the same notions of semilinearity (resp. semialgebraicity).

Example 2.6. Any piecewise linear function is semilinear. For example, any neural network with ReLU activation functions models a semilinear function. Note, however, that, in general, semilinear functions need not be continuous.

The stochastic subgradient method. Let $\mathcal{L} : \mathbb{R}^d \to \mathbb{R}$ be differentiable almost everywhere; this is automatically the case if $\mathcal{L}$ is locally Lipschitz, by Rademacher's theorem (see, e.g., (Evans & Gariepy, 2015)). The *Clarke subdifferential* (Clarke, 1975) of $\mathcal{L}$ at $z \in \mathbb{R}^d$ is

$$\partial\mathcal{L}(z) := \mathsf{ConvHull}\left\{\lim_{z_k \to z} \nabla\mathcal{L}(z_k) : \mathcal{L} \text{ diff. at } \{z_k\}_{k\in\mathbb{N}}\right\}$$

and the point z is *critical* if $0 \in \partial\mathcal{L}(z)$.

In the situation above, one can perform the *stochastic subgradient method* by choosing a learning rate in the form of a sequence $\{a_k \in \mathbb{R}\}_{k\in\mathbb{N}}$, a sequence $\{\zeta_k\}_{k\in\mathbb{N}}$ of real random variables, and an initialization $x_0 \in \mathbb{R}^d$, and defining $\{x_k\}_{k\in\mathbb{N}}$ for $k \geq 1$ recursively by

$$x_{k+1} := x_k - a_k(y_k + \zeta_k), \text{ with } y_k \in \partial\mathcal{L}(x_k).$$

See Assumption B.1 for standard assumptions about the choices involved in the stochastic subgradient method, as taken from Assumption C of (Davis et al., 2020). And see Remark B.5, which addresses satisfying these assumptions.

3 Theoretical contributions

3.1 Stratifying the space of multifiltrations

In this section, we partition the space $\mathsf{Fil}_n(K)$ into finitely many affine regions, which we call cells, and use this to prove Corollary 3.6, which provides an easily verifiable condition for a function with domain $\mathsf{Fil}_n(K)$ to be semilinear. We also show that cells are linearly diffeomorphic to open, convex, semilinear sets of a Euclidean space.

Figure 2 illustrates some of the main concepts introduced in this section.

3.1.1 Cells

In order to define cells, we first define grids, a very simple type of finite poset, and grids obtained from filtrations.

Given $m \in \mathbb{N}$, we let $[m] = \{0 < 1 < \cdots < m-1\}$ denote the corresponding linear order.

Definition 3.1. A *grid* $\mathcal{G}$ is any product poset $\mathcal{G} = \mathcal{G}_1 \times \cdots \times \mathcal{G}_n$, with $\mathcal{G}_i = [m_i]$ for some $m_1, \ldots, m_n \geq 1 \in \mathbb{N}$.

Note that, given a function $f : X \to \mathbb{R}$, there exists a unique triple $(m \in \mathbb{N}, \mathsf{ord}_f : X \to [m], \iota_f : [m] \to \mathbb{R})$ such that ord_f is surjective, ι_f is monotonic and injective, and $f = \iota_f \circ \mathsf{ord}_f$. Indeed, ord_f represents the unique linear preorder of X induced by $\mathsf{Im}(f) \subseteq \mathbb{R}$, so that $m = |\mathsf{Im}(f)|$.

Construction 3.2. Given $f : K \to \mathbb{R}^n$, let $m_i = |f_i(K)|$. Let $\mathsf{grid}_f = [m_1] \times \cdots \times [m_n]$, and define

$$\mathsf{ord}_f \coloneqq \mathsf{ord}_{f_1(K)} \times \cdots \times \mathsf{ord}_{f_n(K)} : K \to \mathsf{grid}_f$$
$$\iota_f \coloneqq \iota_{f_1(K)} \times \cdots \times \iota_{f_n(K)} : \mathsf{grid}_f \to \mathbb{R}^n$$

Note that $f = \iota_f \circ \mathsf{ord}_f : K \to \mathbb{R}^n$.

We can now define the cells of $\mathsf{Fil}_n(K)$, which, informally, are sets of filtrations that induce the same preorder on the simplices of K.

Definition 3.3. The *cell* of $f \in \mathsf{Fil}_n(K)$, denoted $\mathsf{cell}(f) \subseteq \mathsf{Fil}_n(K)$, is the set of all $g \in (\mathbb{R}^n)^K$ such that $\mathsf{grid}_f = \mathsf{grid}_g$ and $\mathsf{ord}_f = \mathsf{ord}_g$.

Remark. Another reasonable definition of cell would be to directly draw from the one-parameter case and declare the cell of a filtration $f : K \to \mathbb{R}^n$ to be the set of filtrations $g : K \to \mathbb{R}^n$ that induce the same preorder on the simplices of K as f does; the difference with our choice being that we require each coordinate $g_i : K \to \mathbb{R}$ to induce the same preorder on the simplices of K as f_i. For the purposes of this remark, let us call this $\mathsf{cell}'(f)$. It can be easily seen that $\mathsf{cell}'(f)$ is a disjoint union of cells in the sense of Definition 3.3. But, importantly, $\mathsf{cell}'(f)$ is typically not convex, or even connected: take K consisting of two points and no other simplices, then $g : K \to \mathbb{R}^2$ mapping the first point to $(0,1)$ and the second to $(1,0)$ is in $\mathsf{cell}'(f)$, where f maps the first point to $(1,0)$ and the second to $(0,1)$; but f and g cannot be connected through a path that stays within $\mathsf{cell}'(f)$. It can also be checked that usual descriptors restricted to $\mathsf{cell}'(f)$ do not behave as nicely as when restricted to $\mathsf{cell}(f)$: for example, the sorted Hilbert decomposition is affine on $\mathsf{cell}(f)$, but typically not on $\mathsf{cell}'(f)$.

We have the following key properties of cells.

Lemma 3.4. *The function $(\mathbb{R}^n)^K \to (\mathbb{R}^K)^n$ mapping g to $(g_1, \ldots, g_n)$ is a linear diffeomorphism. Given $f \in \mathsf{Fil}_n(K)$, this function restricts to a linear diffeomorphism $I_f : \mathsf{cell}(f) \cong \mathsf{cell}(f_1) \times \cdots \times \mathsf{cell}(f_n)$.*

Lemma 3.5. *The set $\mathsf{Fil}_n(K) \subseteq (\mathbb{R}^n)^K$ is semilinear, and, $\mathsf{cell}(f) \subseteq (\mathbb{R}^n)^K$ is semilinear for every $f \in \mathsf{Fil}_n(K)$.*

Corollary 3.6. *Let $\alpha : \mathsf{Fil}_n(K) \to \mathbb{R}^D$ be a function such that $\alpha|_{\mathsf{cell}(f)} : \mathsf{cell}(f) \to \mathbb{R}^D$ is semilinear (resp. semialgebraic) for every $f \in \mathsf{Fil}_n(K)$. Then, α is semilinear (resp. semialgebraic).* □

3.1.2 Grid inclusions and geometry of cells

Definition 3.7. Let $\mathcal{G} = \mathcal{G}_1 \times \cdots \times \mathcal{G}_n$. An *aligned grid inclusion* of $\mathcal{G}$ into $\mathbb{R}^n$ is a monotonic and injective map $\mathcal{G} \to \mathbb{R}^n$ that is a product of monotonic and injective maps $\mathcal{G}_k \to \mathbb{R}$ for $1 \leq k \leq n$.

As an example, for any filtration $f \in \mathsf{Fil}_n(K)$, the map $\iota_f : \mathsf{grid}_f \to \mathbb{R}^n$ is an aligned grid inclusion.

The following construction allows us to identify the space of aligned grid inclusions of a fixed grid $\mathcal{G}$ as a particularly simple subset of a Euclidean space.

Construction 3.8. Given $m \in \mathbb{N}$, define $\mathsf{incl}([m]) \subseteq \mathbb{R}^{[m]}$

$$\mathsf{incl}([m]) \coloneqq \{f : [m] \to \mathbb{R} \mid f(0) < \cdots < f(m-1)\},$$

and, for a grid $\mathcal{G} = \mathcal{G}_1 \times \cdots \times \mathcal{G}_n$, let

$$\mathsf{incl}(\mathcal{G}) \coloneqq \mathsf{incl}(\mathcal{G}_1) \times \cdots \times \mathsf{incl}(\mathcal{G}_n) \subseteq \mathbb{R}^{[m_1]} \times \cdots \times \mathbb{R}^{[m_n]}.$$

Note that, if $\kappa : \mathcal{G} \to \mathbb{R}^n$ is an aligned grid inclusion, then $\kappa = \kappa_1 \times \cdots \times \kappa_n$, and $(\kappa_1, \ldots, \kappa_n) \in \mathsf{incl}(\mathcal{G})$.

Thanks to Construction 3.8, we can, and do, identify aligned grid inclusions of $\mathcal{G}$ into $\mathbb{R}^n$ with $\mathsf{incl}(\mathcal{G})$. In particular, if $f \in \mathsf{Fil}_n(K)$, we write $\iota_f \in \mathsf{incl}(\mathsf{grid}_f)$.

Lemma 3.9. *For any grid $\mathcal{G}$, the set $\mathsf{incl}(\mathcal{G}) \subseteq \mathbb{R}^{[m_1]} \times \cdots \times \mathbb{R}^{[m_n]}$ is open, convex, and semilinear.*

Our goal now is to prove that the map taking a filtration f to its corresponding aligned grid inclusion restricts to a linear diffeomorphism between the cell of f and the space of inclusions $\mathsf{incl}(\mathsf{grid}_f)$.

We start by formally defining this map.

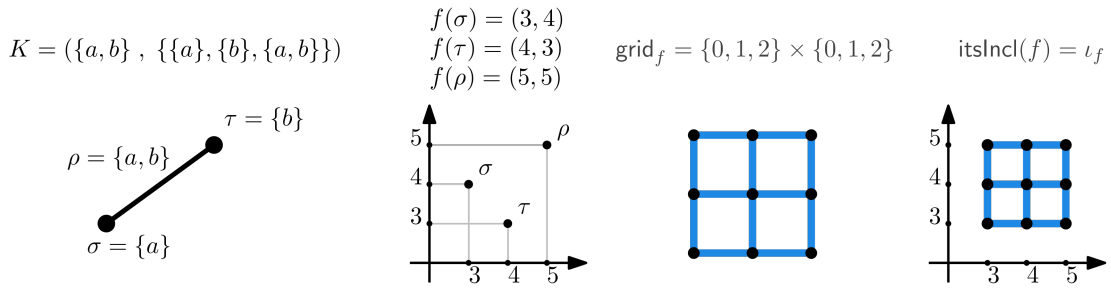


Figure 2. An example of a simplicial complex K, a two-parameter filtration f of K, the grid of f according to Construction 3.2, and the aligned grid inclusion associated with f according to Construction 3.2 and Definition 3.10.

By Definition 3.3, if $g \in \mathsf{cell}(f)$, then $\mathsf{grid}_g = \mathsf{grid}_f$, so that ι_g is an aligned grid inclusion of grid_f into $\mathbb{R}^n$, and thus $\iota_g \in \mathsf{incl}(\mathsf{grid}_f)$. This allows us to define the following.

Definition 3.10. Let $f \in \mathsf{Fil}_n(K)$. Define the function $\mathsf{itsIncl} : \mathsf{cell}(f) \to \mathsf{incl}(\mathsf{ord}_f)$ by mapping $g \in \mathsf{cell}(f)$ to $\iota_g : \mathsf{grid}_f = \mathsf{grid}_g \to \mathbb{R}^n$, seen as an aligned grid inclusion (Construction 3.8).

Definition 3.11. Let $f \in \mathsf{Fil}_1(K)$. A *carrier* for f is a function $C : \mathsf{grid}_f \to K$, such that $\mathsf{ord}_f(C(k)) = k$ for every $k \in \mathsf{grid}_f$. A *carrier* for a multifiltration $f \in \mathsf{Fil}_n(K)$ consists of a family of functions $C = \{C_i : \mathsf{grid}_{f_i} \to K\}_{1 \le i \le n}$, with C_i a carrier of f_i for every $1 \le i \le n$.

Every filtration $f \in \mathsf{Fil}_n(K)$ admits a carrier, since $\mathsf{ord}_{f_i} : K \to \mathsf{ord}_{f_i}$ is surjective for all $1 \le i \le n$, by construction.

Lemma 3.12. *Let $f \in \mathsf{Fil}_n(K)$ and let C be a carrier for f. The function*

$$\prod_{i=1}^{n} \mathsf{cell}(f_i) \xrightarrow{C^*} \prod_{i=1}^{n} \mathsf{incl}(\mathsf{grid}_{f_i}) = \mathsf{incl}(\mathsf{grid}_f)$$

given by mapping $(g_1, \dots g_n)$ to $(g_1 \circ C_1, \dots, g_n \circ C_n)$ is a linear diffeomorphism.

Corollary 3.13. *Let $f \in \mathsf{Fil}_n(K)$ and let C be a carrier for f. Then, $\mathsf{itsIncl} = C^* \circ I_f : \mathsf{cell}(f) \to \mathsf{incl}(\mathsf{grid}_f)$. Thus, $\mathsf{itsIncl} : \mathsf{cell}(f) \to \mathsf{incl}(\mathsf{grid}_f)$ is a linear diffeomorphism between $\mathsf{cell}(f)$ and an open, convex, and semilinear subset of a Euclidean space.*

Note, however, that the dimension of $\mathsf{cell}(f)$ depends on f (more specifically, on grid_f).

3.2 Semilinear descriptors of multifiltrations

In this section we prove Theorem 3.19, which gives sufficient conditions for an invariant to be semilinear as well as an explicit description of the gradient. We instantiate this result to well known invariants in Theorem 1.2.

Definition 3.14. Let $\mathcal{G}$ be a grid. A *$\mathcal{G}$-filtration* of a simplicial complex K consists of a monotonic function $K \to \mathcal{G}$.

$$\begin{tikzcd}
\mathsf{Fil}_n(K) \arrow[rr, "\alpha"] & & \mathbb{R}^D \\
& \mathsf{cell}(f) \arrow[ul, hook] \arrow[r, "\mathsf{itsIncl}\ (\text{Def. } 3.10)", "\cong\ (\text{Cor. } 3.13)"'] \arrow[d, "\cong"', "I_f\ (\text{Lem. } 3.4)"] & \mathsf{incl}(\mathsf{grid}_f) \arrow[u, "\mathsf{push}^{\alpha}_{\mathsf{ord}_f}"', "(\text{Prop. } 3.15)"] \arrow[d, equal] \\
& \prod_{i=1}^{n} \mathsf{cell}(f_i) \arrow[r, "\cong", "C^*\ (\text{Lem. } 3.12)"'] & \prod_{i=1}^{n} \mathsf{incl}(\mathsf{grid}_{f_i})
\end{tikzcd}$$

Figure 3. Factoring a descriptor α through each cell allows us to deal with functions over linear, open subsets of a Euclidean space.

A $\mathcal{G}$-filtration is *componentwise surjective* if $h_i : K \to \mathcal{G}_i$ is surjective for each $1 \le i \le n$.

The set of $\mathcal{G}$-filtrations on K is denoted by $\mathsf{Fil}_{\mathcal{G}}(K)$. As an example, $\mathsf{ord}_f \in \mathsf{Fil}_{\mathsf{grid}_f}(K)$ is a componentwise surjective grid_f-filtration, for every filtration $f \in \mathsf{Fil}_n(K)$.

For an interpretation of the next result, see Remark 3.16.

Proposition 3.15. *Let $\alpha : \mathsf{Fil}_n(K) \to A$. For every grid $\mathcal{G}$ and componentwise surjective $\mathcal{G}$-filtration $h \in \mathsf{Fil}_{\mathcal{G}}(K)$, there exists a unique function $\mathsf{push}^{\alpha}_h : \mathsf{incl}(\mathcal{G}) \to A$ such that, for every $f \in \mathsf{Fil}_n(K)$ with $\mathsf{grid}_f = \mathcal{G}$ and $\mathsf{ord}_f = h$ we have $\alpha|_{\mathsf{cell}(f)} = \mathsf{push}^{\alpha}_{\mathsf{ord}_f} \circ \mathsf{itsIncl} : \mathsf{cell}(f) \to A$.*

Graphically, there is a uniquely determined vertical map making the top rectangle of Figure 3 commute; we now give an interpretation of this map.

Remark 3.16. Let $\alpha : \mathsf{Fil}_n(K) \to A$, and fix a grid $\mathcal{G}$ and componentwise surjective $\mathcal{G}$-filtration $h \in \mathsf{Fil}_{\mathcal{G}}(K)$. The map $\mathsf{push}^{\alpha}_h : \mathsf{incl}(\mathcal{G}) \to A$ is to be interpreted as the map that first computes the invariant α on the discrete, $\mathcal{G}$-filtration h, and then pushes this computation along any given grid inclusion $\kappa \in \mathsf{incl}(\mathcal{G})$. Informally, Proposition 3.15 says that any invariant of multifiltrations is determined by an invariant of discrete filtrations (indexed by

finite grids), and by a map push whose job is to push the invariant of a discrete filtration along grid inclusions.

Definition 3.17. A function $\alpha : \mathsf{Fil}_n(K) \to \mathbb{R}^D$ is *semilinearly determined on grids* if, for each grid $\mathcal{G}$ and componentwise surjective $\mathcal{G}$-filtration $h \in \mathsf{Fil}_{\mathcal{G}}(K)$, the map $\mathsf{push}_h^\alpha : \mathsf{incl}(\mathcal{G}) \to \mathbb{R}^D$ is semilinear.

Remark 3.18. More generally, one can define the notions of a function α being *definably determined on grids* with respect to some o-minimal structure. However, this level of generality is not required to handle multiparameter descriptors we have considered.

The next theorem and proposition follow immediately from Corollaries 3.6 and 3.13 and Lemmas 3.4 and 3.9.

Theorem 3.19. *If $\alpha : \mathsf{Fil}_n(K) \to \mathbb{R}^D$ is semilinearly determined on grids, then it is a semilinear function.* □

Proposition 3.20. *If $\alpha : \mathsf{Fil}_n(K) \to \mathbb{R}^D$ is semilinearly determined on grids, then, for every $f \in \mathsf{Fil}_n(K)$, an element of the Clarke differential of the restriction $\alpha|_{\mathsf{cell}(f)} : \mathsf{cell}(f) \to \mathbb{R}^D$ can be obtained by pulling back an element of the Clarke differential of $\mathsf{push}^\alpha_{\mathsf{ord}_f} : \mathsf{incl}(\mathsf{grid}_f) \to \mathbb{R}^D$, which is a semilinear function with domain an open set of a Euclidean space, along the linear diffeomorphism $C^* \circ I_f$, for C any carrier of f.* □

Remark 3.21. The upshot of Theorem 3.19 and Proposition 3.20 is that, in order to implement subgradient descent for an invariant $\alpha : \mathsf{Fil}_n(K) \to \mathbb{R}^D$, it is enough to provide a map $\mathsf{push}_h^\alpha : \mathsf{incl}(\mathcal{G}) \to \mathbb{R}^D$, for each grid $\mathcal{G}$ and $\mathcal{G}$-filtration $h \in \mathsf{Fil}_{\mathcal{G}}(K)$, for which subgradients are known. Recall that the interpretation of push_h^α is in Remark 3.16.

3.3 Applications

We now describe how to use Propositions D.5 and E.4 to obtain definable and locally Lipschitz objective functions based on the Hilbert decomposition signed measure; analogous constructions work for the rank decomposition signed measure. Similar pipelines can be constructed using the multiparameter persistence landscape, thanks to its semilinearity Proposition D.8 and stability (Vipond, 2020). Recall that OT is the optimal transport distance (A.1).

Example 3.22. Let $\mathcal{DSM}_k(\mathbb{R}^n) \subseteq \mathcal{DSM}(\mathbb{R}^n)$ denote the subset of signed measure of total mass $k \in \mathbb{Z}$ (that is, discrete measures such that the number of positive point masses minus the number of negative point masses is k)[1]. Fix $\nu \in \mathcal{DSM}_k(\mathbb{R}^n)$. Then, the function $E' : \mathcal{DSM}_k(\mathbb{R}^n) \to \mathbb{R}$ given by mapping μ to $\mathsf{OT}(\mu, \nu) \in \mathbb{R}$ is 1-Lipschitz. Since the computation of $\mathsf{OT}(\mu, \nu)$ reduces to computing a minimum (over all permutations of the point masses of $\mu^+ + \nu^-$ and $\nu^+ + \mu^-$) of sums of distances, it is easy to find a representative $E : \mathbb{R}^{\mathsf{indMass}} \to \mathbb{R}$ as in Proposition E.4. And, this map E' is in fact semialgebraic. Combining this with Proposition D.5, we conclude that we can use the distance to a fixed signed measure ν, i.e., the following map, as a topological objective satisfying the assumptions of Thm. 1.1:

$$\mathsf{Fil}_n(K) \longrightarrow \mathbb{R} \ \text{ given by } \ f \longmapsto \mathsf{OT}(\mu^{\mathsf{Hil}}_{H_i(f)}, \nu).$$

Example 3.23. Let $\psi : \mathbb{R}^n \to \mathbb{R}$ be a Lipschitz function. We get a Lipschitz function $\mathcal{DSM}(\mathbb{R}^n) \to \mathbb{R}$ by mapping μ to $\int \psi \, \mathrm{d}\mu \in \mathbb{R}$. Since the computation of $\int \psi \, \mathrm{d}\mu$ reduces to a finite sum indexed by the point masses of μ and weighted by ψ, it is easy to find a representative $E : \mathbb{R}^{\mathsf{indMass}} \to \mathbb{R}$ as in Proposition E.4. If the function ψ is definable over some o-minimal structure, then so is E'. Combining this with Proposition D.5, we conclude that we can use integration with respect to ψ, i.e., the following map, as a topological objective which satisfies assumptions of Theorem 1.1:

$$\mathsf{Fil}_n(K) \longrightarrow \mathbb{R} \ \text{ given by } \ f \longmapsto \int \psi \, \mathrm{d}\mu^{\mathsf{Hil}}_{H_i(f)}.$$

4 Numerical experiments

We first illustrate our framework on a toy point cloud optimization experiment inspired from (Carrière et al., 2021). We then show experiments related more specifically to ML applications with synthetic and real-world data (see also Appendix F, for details). We use the implementation in (Loiseaux & Schreiber, 2024).

Point cloud optimization. This is not an application per se, but rather an illustrative example with the following two main goals: (1) To demonstrate, with an experiment that can be easily assessed visually, that our pipeline works as expected, allowing the user to perform multiparameter homological optimization. (2) To demonstrate that multiparameter homology optimization enables the user to optimize with respect to topological structures that can only be captured by considering the relationship between two or more parameters (in this case metric and density structure). We build upon the point cloud example of (Carrière et al., 2021).

Given a finite set of points X sampled uniformly from the unit square in $\mathbb{R}^2$ (Figure 4a), the objective function aims to maximize the distance between the Hilbert decomposition signed measure of the optimized point cloud and the zero signed measure. Intuitively, this makes the topological content of the optimized point cloud as non-trivial as possible. This is in direct analogy with the loss used in (Carrière et al., 2021), which maximizes the optimal transport distance (referred to as Wasserstein distance there) to the zero persistence diagram. In the one-parameter setting this is achieved by increasing the quantity and "size" (persistence) of the "holes" (cycles), while, in our two-parameter setting, the quantity and size of the holes, as well as the density of the constituent points, play a role in the objective function.

[1] One can see that, for every $f \in \mathsf{Fil}_n(K)$, we have $\mu^{\mathsf{Hil}}_{H_i(f)} \in \mathcal{DSM}_k(\mathbb{R}^n)$ for $k = \dim(H_i(K))$, which only depends on K and not on the specific filtration f.

The Vietoris–Rips filtration VR of X—see Definition 5.2 of (Oudot, 2015)—is used for this purpose, via the maximization of the sum of the lengths of the bars in its barcode[2] in homology degree 1. Up to a constant factor, this can be rewritten as the OT distance between the zero measure and the Hilbert decomposition signed measure of $H_1(\mathrm{VR})$:

$$\mathsf{OT}\left(\mu^{\mathsf{Hil}}_{H_1(\mathrm{VR})}(X)\,,\,0\right). \tag{1}$$

While this approach does succeed in creating many cycles, it has two major drawbacks. (a) As observed in (Carrière et al., 2021) and illustrated in Figure 4c, the topological loss (1) scales with rescaling of the points of X, so loss (1) alone causes the points to diverge to infinity. To mitigate this effect, another purely geometric loss is added in (Carrière et al., 2021), which restricts the points to the unit square. (b) There is no control over the actual sizes of the obtained holes, which end up being arbitrary—see Figure 4d. Here we address these two issues at the same time by using the function-Rips bifiltration $(\mathrm{VR}, \hat{f})$ of X—see Definition 5.1 of (Botnan & Lesnick, 2023)—for some data-dependent kernel density estimator $\hat{f}$, and by maximizing the analog of loss (1) in this two-parameter setting:

$$\mathsf{OT}\left(\mu^{\mathsf{Hil}}_{H_1(\mathrm{VR},\hat{f})}(X)\,,\,0\right). \tag{2}$$

Note that the density estimator $\hat{f}$ is recomputed after each epoch of the optimization. As can be seen from the result (Figure 4b), adding in the density parameter prevents the points from diverging to infinity, the intuitive reason being that the density scales down with rescalings of the points of X. For the same reason, the density parameter exerts some control over the sizes of the cycles produced, inducing larger cycles in lower-density areas.

Another distinguishing feature of the two-parameter metric- and density-aware topological loss is that it preserves some of the characteristics of the density of the initial point cloud. To see this, consider a dataset comprised primarily of points sampled from a large circle, with some extra noise of higher density located around the center (Figure 4e). Using the function-Rips bifiltration and maximizing (2) creates cycles from the noise points in the center, while still maintaining the initial large circle (Figure 4f). In contrast, one-parameter VR optimization treats all points equally, regardless of their density, and therefore destroys the large circle (Figures 4g and 4h). A precise explanation for this phenomenon remains to be determined; we suspect that this behavior is similar to that of mean-shift, in that creating many, or large, holes in high-density areas requires many points in those areas.

Topological dimension reduction. In this experiment, we present an application of multiparameter homological loss differentiation to dimension reduction with autoencoders. Regularizing standard autoencoders with topological losses that constrain the latent space to have the same PH as the input space was one of the first applications of PH-based optimization to appear in the literature (Moor et al., 2020; Doraiswamy et al., 2020; Carrière et al., 2021; Vandaele, Robin and Kang, Bo and Lijffijt, Jefrey and De Bie, Tijl and Saeys, Yvan, 2022). Common topological losses for autoencoders have a regularization term involving a distance between the barcodes of one-parameter VR filtrations:

$$\mathcal{L}(\theta) := \mathsf{OT}\left(\mu^{\mathsf{rk}}_{H_*(\mathrm{VR})}(\tilde{X}(\theta))\,,\,\mu^{\mathsf{rk}}_{H_*(\mathrm{VR})}(X)\right), \tag{3}$$

where θ denotes the parameters of the autoencoder, and X and $\tilde{X}$ denote the input and the learned latent spaces, respectively. This loss does not take density into account, making the quality of the learned latent spaces susceptible to the presence of even mild noise. We compare this to using the following loss, based on the Hilbert decomposition signed measure and a function-Rips two-parameter filtration,

$$\mathcal{L}(\theta) := \mathsf{OT}\left(\mu^{\mathsf{Hil}}_{H_*(\mathrm{VR},\hat{f})}(\tilde{X}(\theta))\,,\,\mu^{\mathsf{Hil}}_{H_*(\mathrm{VR},\hat{f})}(X)\right),$$

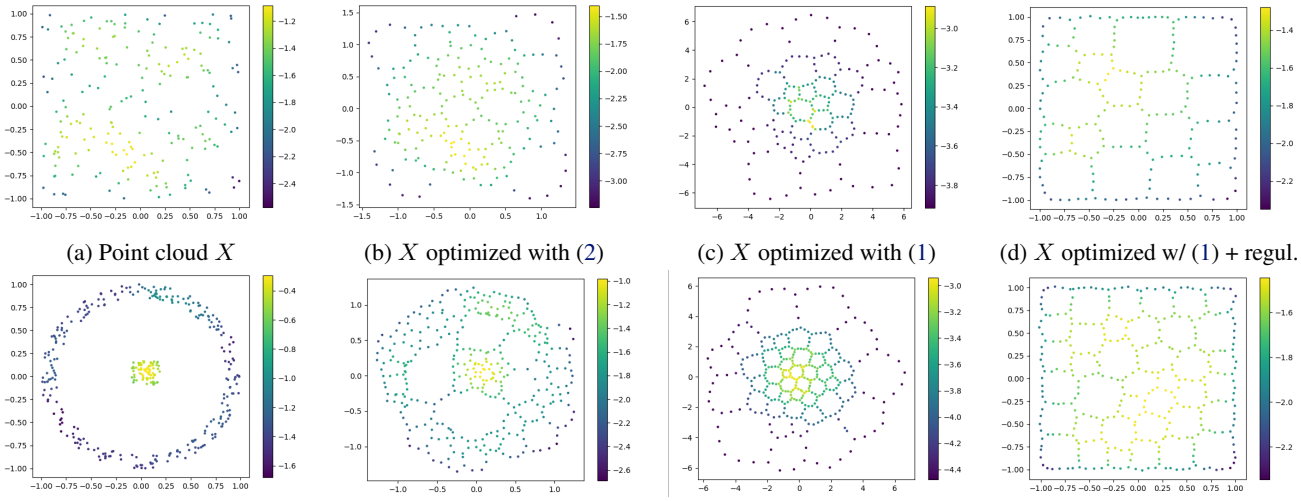
where $\widehat{f}$ is a data-dependent density estimator. This topological loss is an instance of Example 3.22.

The point cloud data consist of two interlaced circles with background noise embedded in $\mathbb{R}^9$ (Figure 7), similar to the data used in (Carrière et al., 2021). See Figure 5, which shows that the proposed multiparameter topological regularization outperforms both no topological regularization, and the one-parameter regularization of Equation (3). See Appendix F.1 for details about this experiment.

Graph classification. In this experiment, we illustrate the use of multiparameter topological optimization for deep learning on graph data, since graph classification is known to be a useful application of topology-based methods (Zhao & Wang, 2019; Hofer et al., 2020b; Carrière et al., 2020; Zhao et al., 2020a; Horn et al., 2022b; Loiseaux et al., 2023b). We use the setup developed in (Horn et al., 2022b), in which node and edge attributes are learned at the same time that a graph neural network classifier is trained, with the attributes being computed using topological descriptors. The authors of (Horn et al., 2022b) study the performance of using k one-parameter descriptors, using k different node attributes. We compare the use of k, independent, one-parameter topological descriptors, against the use of a single k-parameter topological descriptor. Since attributes are learned, the descriptors used must be differentiable, and this is where the application of our theory lies. Details are in Appendix F.2.

Table 1 shows classification performances of a graph neural network architectures complemented with either no topology (*-NOTOP), one-parameter PH as in (Horn et al., 2022b) (*-OPTOP), and multiparameter PH (*-MPTOP).

[2] I.e., one-parameter rank decomposition measure (App. A.2).



(a) Point cloud X (b) X optimized with (2) (c) X optimized with (1) (d) X optimized w/ (1) + regul.

(e) Point cloud Y (f) Y optimized with (2) (g) Y optimized with (1) (h) Y optimized w/ (1) + regul.

Figure 4. Optimizing the holes of point clouds. The colors indicate the log-values of the density estimator.

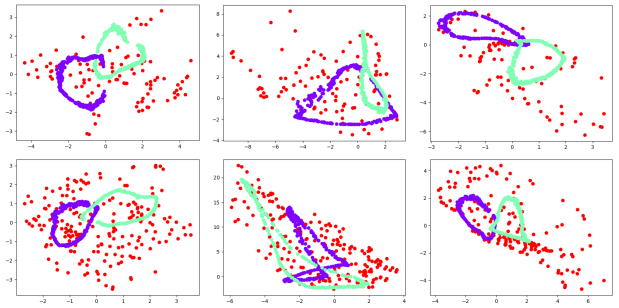


Figure 5. Rows correspond to datasets (top has less background noise), and columns correspond to no topological regularization, one-parameter regularization, and multiparameter regularization. One-parameter regularization is very susceptible to noise points, while no topological regularization can fail to close the circles and preserve topology. Interestingly, no topological regularization behaves better with many noise points, possibly due to the metric loss having more distances to work with. Multiparameter topological regularization ensures the preservation of topology in both cases.

Model	ENZYMES	IMDB-B	IMDB-M	MUTAG
GCN-NOTOP	30.3±8.1	73.2±6.4	44.9±7.6	**87.2±5.6**
GCN-OPTOP	28.8±7.5	75.2±5.6	**51.2±4.4**	84.1±8.9
GCN-MPTOP	**39.0±10.1**	**78.4±5.1**	51.1±3.5	85.1±7.7
GIN-NOTOP	**47.0±12.9**	71.2±5.4	47.1±2.9	87.2±8.0
GIN-OPTOP	45.3±11.8	**75.0±2.7**	47.5±5.0	**88.3±8.9**
GIN-MPTOP	46.5±11.2	71.3±5.1	**48.5±4.2**	87.2±6.1
GraphResNet-NOTOP	42.8±11.0	**75.3±5.3**	49.4±4.3	88.8±5.2
GraphResNet-OPTOP	39.5±12.2	68.1±8.2	40.7±3.5	87.8±4.3
GraphResNet-MPTOP	**44.3±9.8**	69.4±5.8	**50.1±4.4**	**89.3±6.1**
GraphDenseNet-NOTOP	43.2±10.4	50.3±5.9	33.1±2.7	88.8±5.2
GraphDenseNet-OPTOP	47.3±12.3	50.0±7.1	32.7±4.2	86.2±8.3
GraphDenseNet-MPTOP	**48.0±11.4**	**52.2±7.7**	**34.1±3.1**	**92.6±5.1**

Table 1. Bold denotes best score and underline best between one- and multiparameter topology. As one can see, it is usually better to use differentiable multiparameter topological descriptors.

Our goal is *not* to obtain state-of-the-art results, but rather show a positive improvement when using multiparameter homological descriptors instead of one-parameter ones.

5 Conclusions

In order to address the fact that there is no unique homological descriptor of multifiltrations, we introduce a theoretical framework for studying the differentiability and gradient descent convergence of general topology-based losses of multiparameter filtrations. Our main theoretical results give conditions ensuring differentiability and convergence of the stochastic subgradient method, and show that this applies to topological descriptors of very different flavors, demonstrating the flexibility of our approach. Our numerical experiments show positive improvements with respect to one-parameter homological optimization.

Limitations and future work. An inherent limitation of multiparameter persistent homology is that it is generally computationally demanding. This can be dealt with in practice by subsampling or other sparsification techniques (Alonso et al., 2023). To complement this, one could adapt, to the multiparameter case, the optimization with big steps of (Nigmetov & Morozov, 2024), to speed up optimization.

It remains a possibility that our theory does not apply to topological descriptors to be introduced in the future; we believe that suitable generalizations (e.g., descriptors that are definably determined on grids, as in Remark 3.18), will at the very least encompass a large class of descriptors.

On the applications side, future work includes regularizing NNs (Chen et al., 2019) with multiparameter persistence.

Acknowledgements

L.S. was partially supported by EPSRC grant "New Approaches to Data Science: Application Driven Topological Data Analysis", EP/R018472/1. For the purpose of Open Access, the authors have applied a CC BY public copyright licence to any Author Accepted Manuscript (AAM) version arising from this submission. D.L. was supported by ANR grant "3IA Côte d'Azur", ANR-19-P3IA-0002. M.C. was partially supported by ANR grant "TopModel", ANR-23-CE23-0014.

Impact statement

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

References

Alonso, Á. J., Kerber, M., and Pritam, S. Filtration-domination in bifiltered graphs. In *2023 Proceedings of the Symposium on Algorithm Engineering and Experiments (ALENEX)*, pp. 27–38. SIAM, 2023.

Bauer, U. and Scoccola, L. Generic multi-parameter persistence modules are nearly indecomposable. *arXiv preprint arXiv:2211.15306*, 2023.

Beecher, A. Upper bounds for Betti numbers of multigraded modules. *J. Algebra*, 316(1):453–458, 2007. ISSN 0021-8693,1090-266X. doi: 10.1016/j.jalgebra.2007.03.039. URL https://doi.org/10.1016/j.jalgebra.2007.03.039.

Billingsley, P. *Probability and measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, third edition, 1995. ISBN 0-471-00710-2. A Wiley-Interscience Publication.

Botnan, M. B. and Lesnick, M. An introduction to multiparameter persistence. *Proceedings of the 2020 International Conference on Representations of Algebras (to appear). arXiv preprint arXiv:2203.14289*, 2023.

Botnan, M. B., Oppermann, S., and Oudot, S. Signed barcodes for multi-parameter persistence via rank decompositions and rank-exact resolutions, 2022a.

Botnan, M. B., Oppermann, S., Oudot, S., and Scoccola, L. On the bottleneck stability of rank decompositions of multi-parameter persistence modules. *arXiv preprint arXiv:2208.00300*, 2022b.

Bubenik, P. and Elchesen, A. Virtual persistence diagrams, signed measures, Wasserstein distances, and Banach spaces. *J. Appl. Comput. Topol.*, 6(4):429–474, 2022. ISSN 2367-1726,2367-1734. doi: 10.1007/s41468-022-00091-9. URL https://doi.org/10.1007/s41468-022-00091-9.

Carlsson, G. and Zomorodian, A. The theory of multidimensional persistence. *Discrete Comput. Geom.*, 42(1):71–93, 2009. ISSN 0179-5376,1432-0444. doi: 10.1007/s00454-009-9176-0. URL https://doi.org/10.1007/s00454-009-9176-0.

Carrière, M. and Blumberg, A. Multiparameter persistence image for topological machine learning. In *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*, pp. 22432–22444. Curran Associates, Inc., 2020.

Carrière, M., Chazal, F., Ike, Y., Lacombe, T., Royer, M., and Umeda, Y. PersLay: a neural network layer for persistence diagrams and new graph topological signatures. In *23rd International Conference on Artificial Intelligence and Statistics (AISTATS 2020)*, pp. 2786–2796. PMLR, 2020.

Carrière, M., Chazal, F., Glisse, M., Ike, Y., Kannan, H., and Umeda, Y. Optimizing persistent homology based functions. In *38th International Conference on Machine Learning (ICML 2021)*, volume 139, pp. 1294–1303. PMLR, 2021.

Chen, C., Ni, X., Bai, Q., and Wang, Y. A topological regularizer for classifiers via persistent homology. In Chaudhuri, K. and Sugiyama, M. (eds.), *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pp. 2573–2582. PMLR, 16–18 Apr 2019. URL https://proceedings.mlr.press/v89/chen19g.html.

Chen, Y., Segovia-Dominguez, I., Akcora, C. G., Zhen, Z., Kantarcioglu, M., Gel, Y., and Coskunuzer, B. EMP: Effective multidimensional persistence for graph representation learning. In *The Second Learning on Graphs Conference*, 2023. URL https://openreview.net/forum?id=WScCJnX4ek.

Clarke, F. H. Generalized gradients and applications. *Trans. Amer. Math. Soc.*, 205:247–262, 1975. ISSN 0002-9947,1088-6850. doi: 10.2307/1997202. URL https://doi.org/10.2307/1997202.

Cohen-Steiner, D., Edelsbrunner, H., and Harer, J. Stability of persistence diagrams. *Discrete & Computational Geometry*, 37(1):103–120, 2007.

Corbet, R., Fugacci, U., Kerber, M., Landi, C., and Wang, B. A kernel for multi-parameter persistent homology. *Computers & Graphics: X*, 2:100005, 2019.

Davis, D., Drusvyatskiy, D., Kakade, S., and Lee, J. D. Stochastic subgradient method converges on tame functions. *Found. Comput. Math.*, 20(1):119–154, 2020. ISSN 1615-3375,1615-3383. doi: 10.1007/s10208-018-09409-5. URL https://doi.org/10.1007/s10208-018-09409-5.

Demir, A., Coskunuzer, B., Gel, Y., Segovia-Dominguez, I., Chen, Y., and Kiziltan, B. ToDD: Topological compound fingerprinting in computer-aided drug discovery. In Oh, A. H., Agarwal, A., Belgrave, D., and Cho, K. (eds.), *Advances in Neural Information Processing Systems*, 2022. URL https://openreview.net/forum?id=8hs7qlWcnGs.

Divol, V. and Lacombe, T. Understanding the topology and the geometry of the space of persistence diagrams via optimal partial transport. *J. Appl. Comput. Topol.*, 5(1):1–53, 2021. ISSN 2367-1726. doi: 10.1007/s41468-020-00061-z. URL https://doi-org.ezproxy.neu.edu/10.1007/s41468-020-00061-z.

Doraiswamy, H., Tierny, J., Silva, P., Nonato, L. G., and Silva, C. Topomap: A 0-dimensional homology preserving projection of high-dimensional data. *IEEE Transactions on Visualization and Computer Graphics*, 27(2): 561–571, 2020.

Edelsbrunner, H. and Harer, J. Persistent homology—a survey. In *Surveys on discrete and computational geometry*, volume 453 of *Contemp. Math.*, pp. 257–282. Amer. Math. Soc., Providence, RI, 2008. ISBN 978-0-8218-4239-3. doi: 10.1090/conm/453/08802. URL https://doi.org/10.1090/conm/453/08802.

Edelsbrunner, H., Letscher, D., and Zomorodian, A. Topological persistence and simplification. *Discrete and Computational Geometry*, 28:511–533, 2002.

Evans, L. C. and Gariepy, R. F. *Measure theory and fine properties of functions*. Textbooks in Mathematics. CRC Press, Boca Raton, FL, revised edition, 2015. ISBN 978-1-4822-4238-6.

Figalli, A. and Gigli, N. A new transportation distance between non-negative measures, with applications to gradients flows with Dirichlet boundary conditions. *J. Math. Pures Appl. (9)*, 94(2):107–130, 2010. ISSN 0021-7824,1776-3371. doi: 10.1016/j.matpur.2009.11.005. URL https://doi.org/10.1016/j.matpur.2009.11.005.

Fomenko, A. *Visual geometry and topology*. Springer-Verlag, Berlin, 1994. ISBN 3-540-53361-3. doi: 10.1007/978-3-642-76235-2. URL https://doi-org.ezproxy.neu.edu/10.1007/978-3-642-76235-2. Translated from the Russian by Marianna V. Tsaplina.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

Hensel, F., Moor, M., and Rieck, B. A survey of topological machine learning methods. *Frontiers in Artificial Intelligence*, 4, 2021. ISSN 2624-8212. doi: 10.3389/frai.2021.681108. URL https://www.frontiersin.org/articles/10.3389/frai.2021.681108.

Hofer, C., Kwitt, R., Niethammer, M., and Uhl, A. Deep learning with topological signatures. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, pp. 1633–1643, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964.

Hofer, C., Kwitt, R., Niethammer, M., and Dixit, M. Connectivity-optimized representation learning via persistent homology. In Chaudhuri, K. and Salakhutdinov, R. (eds.), *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 2751–2760. PMLR, 09–15 Jun 2019. URL https://proceedings.mlr.press/v97/hofer19a.html.

Hofer, C., Graf, F., Niethammer, M., and Kwitt, R. Topologically densified distributions. In III, H. D. and Singh, A. (eds.), *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 4304–4313. PMLR, 13–18 Jul 2020a. URL https://proceedings.mlr.press/v119/hofer20a.html.

Hofer, C., Graf, F., Rieck, B., Niethammer, M., and Kwitt, R. Graph filtration learning. In III, H. D. and Singh, A. (eds.), *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 4314–4323. PMLR, 13–18 Jul 2020b. URL https://proceedings.mlr.press/v119/hofer20b.html.

Horn, M., Brouwer, E. D., Moor, M., Moreau, Y., Rieck, B., and Borgwardt, K. Topological graph neural networks. In *International Conference on Learning Representations*, 2022a. URL https://openreview.net/forum?id=oxxUMeFwEHd.

Horn, M., de Brouwer, E., Moor, M., Rieck, B., and Borgwardt, K. Topological Graph Neural Networks. In *10th International Conference on Learning Representations (ICLR 2022)*. OpenReviews.net, 2022b.

Huang, G., Liu, Z., van der Maaten, L., and Weinberger, K. Q. Densely connected convolutional networks. In *CVPR*, pp. 2261–2269. IEEE Computer Society, 2017.

Kim, K., Kim, J., Zaheer, M., Kim, J., Chazal, F., and Wasserman, L. Pllay: Efficient topological layer based on persistence landscapes. In *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*, volume 33, pp. 15965–15977. Curran Associates, Inc., 2020.

Kipf, T. N. and Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. In *Proceedings of the 5th International Conference on Learning Representations*, ICLR '17, 2017.

Leygonie, J., Oudot, S., and Tillmann, U. A framework for differential calculus on persistence barcodes. *Found. Comput. Math.*, 22(4):1069–1131, 2022. ISSN 1615-3375,1615-3383. doi: 10.1007/s10208-021-09522-y. URL https://doi.org/10.1007/s10208-021-09522-y.

Leygonie, J., Carrière, M., Lacombe, T., and Oudot, S. A gradient sampling algorithm for stratified maps with applications to topological data analysis. *Math. Program.*, 202(1-2):199–239, 2023. ISSN 0025-5610,1436-4646. doi: 10.1007/s10107-023-01931-x. URL https://doi.org/10.1007/s10107-023-01931-x.

Loiseaux, D. and Schreiber, H. multipers: Multiparameter persistence for machine learning. https://github.com/DavidLapous/multipers, 2024.

Loiseaux, D., Carrière, M., and Blumberg, A. A Framework for Fast and Stable Representations of Multiparameter Persistent Homology Decompositions. In *Advances in Neural Information Processing Systems 36 (NeurIPS 2023)*. Curran Associates, Inc., 2023a.

Loiseaux, D., Scoccola, L., Carrière, M., Botnan, M., and Oudot, S. Stable Vectorization of Multiparameter Persistent Homology using Signed Barcodes as Measures. In *Advances in Neural Information Processing Systems 36 (NeurIPS 2023)*. Curran Associates, Inc., 2023b.

Moor, M., Horn, M., Rieck, B., and Borgwardt, K. Topological autoencoders. In *37th International Conference on Machine Learning (ICML 2020)*, volume 119, pp. 7045–7054. PMLR, 2020.

Mukherjee, S., Shreyas, S. N., Xin, C., Oudot, S., and Dey, T. K. D-GRIL: End-to-End Topological Learning with 2-parameter Persistence. 2024.

Nigmetov, A. and Morozov, D. Topological optimization with big steps. *Discrete & Computational Geometry*, pp. 1–35, 2024.

Oudot, S. and Scoccola, L. On the stability of multigraded betti numbers and hilbert functions. *SIAM Journal on Applied Algebra and Geometry*, 8(1):54–88, 2024. doi: 10.1137/22M1489150. URL https://doi.org/10.1137/22M1489150.

Oudot, S. Y. *Persistence theory: from quiver representations to data analysis*, volume 209 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2015. ISBN 978-1-4704-2545-6. doi: 10.1090/surv/209. URL https://doi.org/10.1090/surv/209.

Poulenard, A., Skraba, P., and Ovsjanikov, M. Topological function optimization for continuous shape matching. *Computer Graphics Forum*, 37(5):13–25, 2018.

van den Dries, L. *Tame topology and o-minimal structures*, volume 248 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 1998. ISBN 0-521-59838-9. doi: 10.1017/CBO9780511525919. URL https://doi.org/10.1017/CBO9780511525919.

van den Dries, L. and Miller, C. Geometric categories and o-minimal structures. *Duke Math. J.*, 84(2):497–540, 1996. ISSN 0012-7094,1547-7398. doi: 10.1215/S0012-7094-96-08416-1. URL https://doi.org/10.1215/S0012-7094-96-08416-1.

Vandaele, Robin and Kang, Bo and Lijffijt, Jefrey and De Bie, Tijl and Saeys, Yvan. Topologically regularized data embeddings. In *10th International Conference on Learning Representations (ICLR 2022)*. OpenReviews.net, 2022.

Vipond, O. Multiparameter persistence landscapes. *Journal of Machine Learning Research*, 21(61):1–38, 2020.

Wilkie, A. J. Model completeness results for expansions of the ordered field of real numbers by restricted Pfaffian functions and the exponential function. *J. Amer. Math. Soc.*, 9(4):1051–1094, 1996. ISSN 0894-0347,1088-6834. doi: 10.1090/S0894-0347-96-00216-0. URL https://doi.org/10.1090/S0894-0347-96-00216-0.

Xin, C., Mukherjee, S., Samaga, S., and Dey, T. GRIL: A 2-parameter Persistence Based Vectorization for Machine Learning. In *2nd Annual Workshop on Topology, Algebra, and Geometry in Machine Learning*. OpenReviews.net, 2023.

Xu, K., Hu, W., Leskovec, J., and Jegelka, S. How powerful are graph neural networks? In *7th International Conference on Learning Representations (ICLR 2019)*. OpenReviews.net, 2019.

Zhao, Q. and Wang, Y. Learning metrics for persistence-based summaries and applications for graph classification. In *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)*, pp. 9855–9866. Curran Associates, Inc., 2019.

Zhao, Q., Ye, Z., Chen, C., and Wang, Y. Persistence enhanced graph neural network. In *23rd International Conference on Artificial Intelligence and Statistics (AISTATS 2020)*, pp. 2896–2906. PMLR, 2020a.

Zhao, Q., Ye, Z., Chen, C., and Wang, Y. Persistence enhanced graph neural network. In Chiappa, S. and Calandra, R. (eds.), *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pp. 2896–2906. PMLR, 26–28 Aug 2020b. URL https://proceedings.mlr.press/v108/zhao20d.html.

Zomorodian, A. and Carlsson, G. Computing persistent homology. *Discrete & Computational Geometry*, 33(2): 249–274, 2005.

A More background

A.1 Partial optimal transport

A.1.1 Definition

We define optimal transport between signed measures on a metric spaces, and partial optimal transport between signed measures on a metric pair. This allows one to model the scenario in which one considers signed measures on a space M and optimal transport plans that are allowed to "throw mass away" in a certain subset $Z \subseteq M$. Partial optimal transport is relevant in topological data analysis (Divol & Lacombe, 2021) and in optimal transport problems in general (Figalli & Gigli, 2010). For more context, see Appendix A.1.2.

Let M be an *extended metric space*, that is, a set together with a function $d_M : M \times M \to [0, \infty]$ that is symmetric, which satisfies the triangle inequality, and such that $d_M(x, y) = 0$ if and only if $x = y \in M$.

The set $\mathcal{DM}(M)$ of discrete measures on M admits an extended metric, called the *optimal transport distance* and denoted OT, which we now define. Given $\mu, \nu \in \mathcal{DM}(M)$, let

$$\mathsf{assign}(\mu, \nu) := \left\{ \left(I \text{ fin. set}, \begin{smallmatrix} \beta : I \to M \\ \gamma : I \to M \end{smallmatrix} \right) : \begin{smallmatrix} \mu = \sum_{i \in I} \delta_{\beta(i)} \\ \nu = \sum_{i \in I} \delta_{\gamma(i)} \end{smallmatrix} \right\},$$

and for every $a = (I, \beta, \gamma) \in \mathsf{assign}(\mu, \nu)$, define $\mathsf{cost}(a) := \sum_{i \in I} d_M(\beta(i), \gamma(i))$. Then, the set $\mathcal{DM}(M)$ together with the function

$$\mathsf{OT}(\mu, \nu) := \inf\{\mathsf{cost}(a) : a \in \mathsf{assign}(\mu, \nu)\},$$

with the convention that the infimum over an empty set is $+\infty$, is an extended metric space.

The optimal transport distance extends to $\mathcal{DSM}(M)$, the set of discrete signed measures on M, by reduction to the positive case: define $\mathsf{OT}(\mu, \nu) := \mathsf{OT}(\mu^+ + \nu^-, \nu^+ + \mu^-)$, where the signs $+$ and $-$ in superscripts indicate the positive and negative parts of the signed measures according to their Jordan decomposition.

Then, discrete signed measures on a metric pair are defined as follows.

Definition A.1. If $Z \subseteq M$ is any subset, the set of *discrete signed measures* on the pair (M, Z) is $\mathcal{DSM}(M, Z) := \mathcal{DSM}(M \setminus Z)$.

It may seem from Definition A.1 that the subset Z is hardly playing any role; this is not so, since the partial optimal transport distance (Definition A.2) uses Z in a crucial way.

One can extend the optimal transport distance on discrete signed measures to the partial optimal transport distance on the set $\mathcal{DSM}(M, Z)$ of discrete signed measures on (M, Z) that, informally, allows one to "throw mass away" in the subset Z. This is defined formally as follows.

Definition A.2. Let $\mu, \nu \in \mathcal{DSM}(M, Z)$, their *partial optimal transport distance* is

$$\mathsf{OT}(\mu, \nu) := \inf \left\{ \mathsf{OT}(\overline{\mu}, \overline{\nu}) \ : \ \overline{\mu}, \ \overline{\nu} \in \mathcal{DSM}(M), \ \ \mu = \overline{\mu}|_{M \setminus Z}, \ \ \nu = \overline{\nu}|_{M \setminus Z} \right\}.$$

Then, the set $\mathcal{DSM}(M, Z)$ endowed with the partial optimal transport distance is an extended metric space (see Appendix A.1.2).

We use the same notation for the optimal transport distance and the partial optimal transport distance, as these are defined over two different spaces.

A.1.2 Context

Note that, depending on the context, optimal transport distances are named after different people, including Kantorovich, Monge, Rubinstein, and Wasserstein.

Since in this paper we deal with discrete measures, our optimal transport distances take a particularly simple form, which reduces its computation to an assignment problem; we refer the reader to (Bubenik & Elchesen, 2022) for a detailed exposition—from the lens of topological data analysis—of the concepts introduced here; in the cited paper, the measures we consider are called virtual persistence diagrams (Definition 4.10 of (Bubenik & Elchesen, 2022)). Then, Corollary 4.9

of (Bubenik & Elchesen, 2022) implies that our partial optimal transport distance (Definition A.2) does in fact endow $\mathcal{DSM}(M, Z)$ with an extended metric.

Partial optimal transport between signed measures is not as develop as its positive (i.e., unsigned) counterpart, but note that Proposition 5.19 of (Bubenik & Elchesen, 2022) gives very general conditions under which their notion of partial optimal transport for positive measures (which, in the case of discrete measures restricts to ours) coincides with that of (Figalli & Gigli, 2010).

A.2 The rank decomposition signed measure

The rank decomposition signed measure is a generalization of the well-known one-parameter barcode (also known as persistence diagram); and instead of being a measure on the space of intervals (or bars) in $\mathbb{R}$, it is a signed measure on the space of higher dimensional bars, which we now define. If $n \geq 1 \in \mathbb{N}$, let

$$\mathsf{bars}_n = \{(r, s) \in (\mathbb{R}^n \cup \{\infty\})^2 : r \leq s\} \subseteq (\mathbb{R}^n \cup \{\infty\})^2 .$$

As in the one-parameter case, short bars are bars close to the diagonal, defined as $\Delta \coloneqq \{(r, r)\} \subseteq \mathsf{bars}_n$.

Definition A.3. Let $f \in \mathsf{Fil}_n(K)$ and let $i \in \mathbb{N}$. The ith *rank decomposition signed measure* is the unique signed measure $\mu^{\mathsf{rk}}_{H_i(f)} \in \mathcal{DSM}(\mathsf{bars}_n, \Delta)$ such that, for all $r \leq s \in \mathbb{R}^n \cup \{\infty\}$,

$$\mathsf{rk}\left(H_i(f)(r) \xrightarrow{\varphi^f_{r,s}} H_i(f)(s)\right) = \mu^{\mathsf{rk}}_{H_i(f)}\left(C(r, s)\right),$$

where $C(r, s) = \{(r', s') \in \mathsf{bars}_n : r' \leq r, s' \not\leq s\}$.

Existence and uniqueness follows from Corollary 5.6 of (Botnan et al., 2022a).

In order to make bars_n into an (extended) metric space, we endow $\mathbb{R}^n$ and $(\mathbb{R}^n \cup \{\infty\})^2$ with the extended distance induced by $\| - \|_\infty$, with the convention that $|\infty - x| = |x - \infty| = \infty$ if $x \neq \infty$ and $|\infty - \infty| = 0$.

In particular, in the one-parameter case, $\mathcal{DSM}(\mathsf{bars}_1, \Delta)$ is the usual set of one-parameter persistence barcodes, and the descriptor μ^{rk} is the one-parameter barcode.

A.3 Isomorphism invariance

Let K and K' be simplicial complexes with underlying sets X and X', respectively. An *isomorphism* between K and K' consists of a bijection $\psi : X \to X'$ such that $\sigma = \{x_0, \ldots, x_i\} \in K$ if and only if $\psi(\sigma) \coloneqq \{\psi(x_0), \ldots, \psi(x_i)\} \in K'$.

Example A.4. Two graphs are isomorphic as graphs if and only if they are isomorphic as simplicial complexes.

An *isomorphism* between filtrations $f \in \mathsf{Fil}_n(K)$ and $g \in \mathsf{Fil}_n(K')$ consists of an isomorphism of simplicial complexes $\psi : K \to K'$ such that $g \circ \psi = f$.

Definition A.5. A family of descriptors $\{\alpha^K : \mathsf{Fil}_n(K) \to A\}_K$ indexed by all finite simplicial complexes K is *isomorphism invariant* if $\alpha(f) = \alpha(g) \in A$ whenever f and g are isomorphic filtrations.

Since homology is isomorphism invariant, it follows that all descriptors that only rely on homology are also isomorphism invariants. In particular, we have the following.

Proposition A.6. *The Hilbert function, rank invariant, Hilbert decomposition signed measure, and rank decomposition signed measure are isomorphism invariant.* □

B Convergence of the stochastic subgradient method with topological descriptors

B.1 Details about the stochastic subgradient method

Assumption B.1. (a) The sequence $\{a_k\}_{k\in\mathbb{N}}$ is such that $a_k \geq 0$ for all $k \in \mathbb{N}$, $\sum_k a_k = \infty$, and $\sum_k a_k^2 < \infty$.

(b) Almost surely, $\sup_k \|x_k\| < \infty$.

(c) The sequence of random variables $\{\zeta_k\}_{k\in\mathbb{N}}$ has the property that there exists a function $p : \mathbb{R}^d \to [0, \infty)$, bounded on bounded sets, such that

$$\mathbb{E}[\zeta_k|\mathcal{F}_k] = 0 \text{ and } \mathbb{E}[\|\zeta_k\|^2|\mathcal{F}_k] < p(x_k),$$

almost surely, where $\mathcal{F}_k = \sigma(x_j, y_j, \zeta_j)_{j<k}$ is the increasing sequence of σ-algebras generated by x, y, and ζ up to k.

We comment on how to satisfy the Assumption B.1 in Remark B.5

The following is a simple adaptation of a main result from (Davis et al., 2020).

Proposition B.2. *Assume given a simplicial complex K, a number of parameters $n \geq 1 \in \mathbb{N}$, and functions*

$$\mathbb{R}^d \xrightarrow{\Phi} \mathsf{Fil}_n(K) \xrightarrow{\alpha} \mathbb{R}^D \xrightarrow{E} \mathbb{R}.$$

Assume the following:

1. *The loss function $\mathcal{L} := \Phi \circ E \circ \alpha : \mathbb{R}^d \to \mathbb{R}$ is locally Lipschitz.*
2. *The function α is semilinear.*
3. *The functions Φ and E are definable over a common o-minimal structure.*

Under Assumption B.1, almost surely, every limit point of the iterates of the stochastic subgradient method is critical for the loss $\mathcal{L}$, and the sequence of loss values converges.

Proof. We use Corollary 5.9, part (stochastic subgradient method), of (Davis et al., 2020). In the notation of that result, we are considering $f = \mathcal{L} = E \circ \alpha \circ \Phi$, which is locally Lipschitz by assumption (1). By assumption (3), the functions Φ and E are definable over a common o-minimal structure $\mathcal{O}$, and by assumption (2), it follows that $\mathcal{L} = E \circ \alpha \circ \Phi$ is definable over $\mathcal{O}$ (since semilinear functions are definable over all o-minimal structures (van den Dries, 1998)). Since functions definable over an o-minimal structure are Whitney C^p-stratifiable for any $p \geq 1 \in \mathbb{N}$ (van den Dries & Miller, 1996), it follows that $f = \mathcal{L}$ is Whitney C^d-stratifiable. □

B.2 Proof of Theorem 1.1

The statements not involving the optimization pipeline $E \circ \alpha \circ \Phi$ are the contents of Theorem 3.19 and Proposition 3.20. For the convergence of the stochastic subgradient method, we use Proposition B.2, and to satisfy condition (2) we use and Theorem 3.19 and the fact that α is semilinearly determined on grids, by assumption.

B.3 On the assumptions of Theorem 1.1 and possible extensions

We comment on possible uses and generalizations of Proposition B.2 and thus of Theorem 1.1.

Remark B.3. The optimization pipeline considered in Theorem 1.1 typically arises in ML contexts where PH is used as the initial data featurization step. For instance, when the input data are filtered point clouds to be optimized, the map Φ can build the corresponding function-Rips bifiltration, while the map E can be the loss associated to some neural network architecture into which the vectorized invariant is plugged.

Beyond that, Theorem 1.1 can easily be adapted to encompass a larger variety of use cases, using stronger results of (Davis et al., 2020), such as those in Section 6 of (Davis et al., 2020). For instance, one can consider composites $\mathcal{L} \circ F$, where $F \colon \mathbb{R}^p \to \mathbb{R}^d$ is both locally Lipschitz and definable over the same o-minimal structure as $\mathcal{L}$ (for instance, F could be some neural network), which enables for instance the use of PH as an intermediate step (as opposed to the initial step) in the learning pipeline, with back-propagation of the gradients via the chain rule. Alternatively, one can consider sums $\mathcal{L} + \mathcal{L}'$ where $\mathcal{L}'$ is another objective function (possibly non-topological) that is both locally Lipschitz and definable over the same o-minimal structure as $\mathcal{L}$, which enables for instance the use of PH as a topological regularizer. Finally, one can also restrict the domain of Φ, also as regularization.

We now comment on the different conditions of Proposition B.2.

Remark B.4. Assumption B.1 contains technical conditions about the choices made for the stochastic subgradient method. We comment on how these can be met in Remark B.5.

Condition (3) is often easy to meet; for example, by the second main result of (Wilkie, 1996), this holds automatically if the functions consist of combinations of algebraic and exponential functions (see Section 3.3 of (Carrière et al., 2021) for examples for Φ that are particularly relevant in topological optimization).

The theoretical results of this paper are specifically about meeting the conditions involving the topological descriptor α: assumptions (1) and (2).

Finally, we comment on how to satisfy the standard assumptions for the stochastic subgradient method (Assumption B.1).

Remark B.5. Condition (a) is immediate to satisfy, since we have control over the choice of learning rate. Condition (c) is satisfied, for example, by any sequence of random variables $\{\zeta_k\}$ with zero mean, bounded variance, independent, and independent of $\{x_k\}$ and $\{y_k\}$. Although condition (b) seems more difficult to satisfy, as it involves proving something about the behavior of the stochastic subgradient method, it is known that it can be dealt with by adding a suitable regularization (for example, in the form of a non-topological loss $\mathcal{L}'$ as in Remark B.3) forcing the iterates of the stochastic subgradient method to remain in a bounded region, or simply by restricting the domain of Φ; we refer the interested reader to Section 6.1 of (Davis et al., 2020).

C Proofs of Section 3

C.1 Proofs of Section 3.1

Proof of Lemma 3.4. The function $(\mathbb{R}^n)^K \to (\mathbb{R}^K)^n$ in the statement is clearly a linear diffeomorphism, since its inverse is given by mapping $(g_1, \dots, g_n)$ to the function $K \to \mathbb{R}^n$ given by sending σ to $(g_1(\sigma), \dots, g_n(\sigma))$. It follows directly from the definition of cells that this function restricts to a bijection $\mathsf{cell}(f) \cong \mathsf{cell}(f_1) \times \cdots \times \mathsf{cell}(f_n)$, concluding the proof. □

Lemma C.1. *The set* $\{\mathsf{cell}(f) : f \in \mathsf{Fil}_n(K)\}$ *is finite.*

Proof. Given $f \in \mathsf{Fil}_n(K)$, we have $\mathsf{grid}_f = [m_1] \times \cdots \times [m_n]$ with $m_i = |f_i(K)|$, so $m_i \leq |K|$. Thus, there are finitely many product posets grid_f that can be obtained. Moreover, there are finitely many possible maps $K \to \mathsf{grid}_f$ into each of these finitely many posets. □

Proof of Lemma 3.5. The first statement follows from the second and Lemma C.1, since distinct cells are disjoint, the union of all cells is exactly $\mathsf{Fil}_n(K)$, and finite unions of semilinear sets are semilinear. So it remains to show that all cells are semilinear. By Lemma 3.4, it is sufficient to prove it in the one-parameter case, that is for $f \in \mathsf{Fil}_1(K)$.

Given $\sigma, \tau \in K$, define a polynomial $P^{\sigma,\tau} : \mathbb{R}^K \to \mathbb{R}$ as $P^{\sigma,\tau}(h) = h_\sigma - h_\tau$, and a set $S^{\sigma,\tau} \subseteq \mathbb{R}^K$ as follows:

$$S^{\sigma,\tau} = \begin{cases} \{h \in (\mathbb{R}^n)^K : P^{\sigma,\tau}(h) = 0\} & \text{if } f(\sigma) = f(\tau) \\ \{h \in (\mathbb{R}^n)^K : P^{\sigma,\tau}(h) < 0\} & \text{if } f(\sigma) < f(\tau) \\ \{h \in (\mathbb{R}^n)^K : P^{\sigma,\tau}(h) > 0\} & \text{if } f(\sigma) > f(\tau). \end{cases}$$

Note that $\mathsf{ord}_h = \mathsf{ord}_f$ if and only if the preorder on the simplices of K induced by h is equal to that induced by f, and, in turn, this is true if and only if $h \in \bigcap_{\sigma,\tau \in K} S^{\sigma,\tau}$. Thus, $\mathsf{cell}(f) = \bigcap_{\sigma,\tau \in K} S^{\sigma,\tau}$, which is a finite intersection of semilinear sets, and thus semilinear. □

Proof of Lemma 3.9. We prove that $\mathsf{incl}([m]) \subseteq \mathbb{R}^m$ is open, convex, and semilinear for every $m \geq 1 \in \mathbb{N}$; this is sufficient, since $\mathsf{incl}(\mathcal{G}) = \mathsf{incl}(\mathcal{G}_1) \times \cdots \times \mathsf{incl}(\mathcal{G}_n)$ can then be written as a finite intersection of open, convex, and semilinear sets.

The fact that $\mathsf{incl}([m]) \subseteq \mathbb{R}^m$ is open follows directly from its definition, which uses strict inequalities. Convexity follows from the fact that $r_1 < \cdots < r_n$ and $s_1 < \cdots < s_n$ implies $ar_1 + bs_1 < \cdots < ar_n + bs_n$ for any $a, b \geq 0 \in \mathbb{R}$. We reduce the semilinearity statement to Lemma 3.5, as follows. Let K be the simplicial complex on the set $\{0, \dots, m-1\}$ that has one 0-simplex for each element and no higher-dimensional simplices. Consider the filtering function $f \in \mathsf{Fil}_1(K)$ with $f(i) = i$, for $i \in \{0, \dots, m-1\}$. Then, a straightforward check shows that $\mathsf{incl}([m]) = \mathsf{cell}(f)$, so the result follows from Lemma 3.5. □

Proof of Lemma 3.12. Since everything is defined componentwise, it is sufficient to prove this for $f \in \mathsf{Fil}_1(K)$.

The fact that the function is linear is evident from its definition; it thus suffices to prove that it is a bijection. Consider the function $F : \mathsf{incl}(\mathsf{grid}_f) \to \mathsf{cell}(f)$ mapping any injective and monotonic function $\kappa : \mathsf{grid}_f \to \mathbb{R}$ to $\kappa \circ \mathsf{ord}_f$. We prove that F is the inverse of C^*.

Right inverse. If $\kappa \in \mathsf{incl}(\mathsf{grid}_f)$, then $C^*(F(\kappa)) = \kappa \circ \mathsf{ord}_f \circ C = \kappa$, by the fact that C is a carrier for f.

Left inverse. If $g \in \mathsf{cell}(f)$, then $F(C^*(g)) = g \circ C \circ \mathsf{ord}_f = \iota_g \circ \mathsf{ord}_g \circ C \circ \mathsf{ord}_f = \iota_g \circ \mathsf{ord}_f \circ C \circ \mathsf{ord}_f = \iota_g \circ \mathsf{ord}_f = \iota_g \circ \mathsf{ord}_g = g$, using the fact that $g = \iota_g \circ \mathsf{ord}_g$, the fact that $\mathsf{ord}_g = \mathsf{ord}_f$ twice, and the fact that C is a carrier for f. □

C.2 Proofs of Section 3.2

Proof of Proposition 3.15. We start by proving existence. Consider the function $\mathsf{push}_h^\alpha : \mathsf{incl}(\mathcal{G}) \to A$ given by mapping $\kappa \in \mathsf{incl}(\mathcal{G})$ to $\alpha(\kappa \circ h) \in A$. If $f \in \mathsf{Fil}_n(K)$ with $\mathsf{grid}_f = \mathcal{G}$ and $\mathsf{ord}_h = h$ and $g \in \mathsf{cell}(f)$, we have $\mathsf{push}_h^\alpha(\iota_g) = \alpha(\iota_g \circ h) = \alpha(\iota_g \circ \mathsf{ord}_f) = \alpha(\iota_g \circ \mathsf{ord}_g) = \alpha(g)$. This shows existence.

We now show uniqueness. Suppose $p : \mathsf{incl}(\mathcal{G}) \to A$ also satisfies that, if $f \in \mathsf{Fil}_n(K)$ with $\mathsf{grid}_f = \mathcal{G}$ and $\mathsf{ord}_f = h$, then $\alpha = p \circ \mathsf{itsIncl}$. Note that, if $\kappa \in \mathsf{incl}(\mathcal{G})$ and $d = \kappa \circ h$, then since $h_i : K \to \mathcal{G}_i$ is surjective for all $1 \leq i \leq n$, it follows that $\mathsf{grid}_d = \mathcal{G}$ and $\mathsf{ord}_d = h$. Let $f \in \mathsf{Fil}_n(K)$ such that $\mathsf{grid}_f = \mathcal{G}$ and $\mathsf{ord}_f = h$ (note that such a function always exists by taking, e.g., $f = \kappa \circ h$ for any $\kappa \in \mathsf{incl}(\mathcal{G})$). Now, if $\kappa \in \mathsf{incl}(\mathcal{G})$, then $p(\kappa) = p(\iota_{\kappa \circ h}) = \alpha(\kappa \circ h) = \mathsf{push}_h^\alpha(\kappa)$, as required. □

Proof of Corollary 3.13. It's enough to prove that $\mathsf{itsIncl} = C^* \circ I_f : \mathsf{cell}(f) \to \mathsf{incl}(\mathsf{grid}_f)$; the second claim follows then follows directly from Lemmas 3.4, 3.9 and 3.12.

Given $g \in \mathsf{cell}(f)$ and $x \in \mathsf{grid}_f$, we compute as follows

$$\begin{aligned}
\iota_g(x) &= \iota_g(x_1, \dots, x_n) \\
&= \iota_g\big(\mathsf{ord}_{f_1}(C_1(x_1)), \dots, \mathsf{ord}_{f_n}(C_n(x_n))\big) \\
&= \iota_g\big(\mathsf{ord}_{g_1}(C_1(x_1)), \dots, \mathsf{ord}_{g_n}(C_n(x_n))\big) \\
&= \big(\iota_{g_1}(\mathsf{ord}_{g_1}(C_1(x_1))), \dots, \iota_{g_n}(\mathsf{ord}_{g_n}(C_n(x_n)))\big) \\
&= \big(g_1(C_1(x_1)) \dots, g_n(C_n(x_n))\big) \\
&= I_f(g)(C(x)) \\
&= C^*(I_f(g))(x).
\end{aligned}$$

Thus, $\iota_g = C^*(I_f(g)) : \mathsf{grid}_f \to \mathbb{R}^n$, as required. □

D Semilinearity of known invariants

We start by providing a proof of Theorem 1.2, which is a direct consequence of the results in this section.

The sorted Hilbert decomposition is defined in Definition D.4, and the evaluated multiparameter persistence landscape is defined in Definition D.7.

Proof of Theorem 1.2. This is a direct consequence of Proposition D.5 and Proposition D.8. □

D.1 Semilinearity of Hilbert decomposition signed measure

We do the case of the Hilbert decomposition signed measure, and thus that of discrete signed measures on $\mathbb{R}^n$; the case of the rank decomposition signed measure is only slightly more verbose, but not conceptually harder.

In order to prove semilinearity, we must represent the Hilbert decomposition signed measures as elements of a finite dimensional vector space.

For notational clarity, fix $n \geq 1 \in \mathbb{N}$, $i \in \mathbb{N}$, and a simplicial complex K.

In Appendix D.1.1, we prove a technical result stating that the Hilbert decomposition signed measure has a constant number of positive and negative masses on each cell. In Appendix D.1.2, we prove the semilinearity of the sorted Hilbert decomposition.

D.1.1 Constant number of point masses

It is convenient to also consider discrete signed measures on discrete sets, such as grids. In this case, we do not define optimal transport distances, since we do not make use of these.

Lemma D.1. *Let $i \in \mathbb{N}$. Let $\mathcal{G} = \mathcal{G}_1 \times \cdots \times \mathcal{G}_n$ be a finite grid and let $h \in \mathsf{Fil}_{\mathcal{G}}(K)$. There exists a signed measure $\mu^{\mathsf{Hil}}_{H_i(h)} \in \mathcal{DSM}(\mathcal{G})$ such that, for every grid inclusion $\kappa \in \mathsf{incl}(\mathcal{G})$ we have*

$$\mu^{\mathsf{Hil}}_{H_i(\kappa \circ h)} = \kappa_{\#}\, \mu^{\mathsf{Hil}}_{H_i(h)} \ \in\ \mathcal{DSM}(\mathbb{R}^n),$$

where $\kappa_{\#} : \mathcal{DSM}(\mathcal{G}) \to \mathcal{DSM}(\mathbb{R}^n)$ denotes the push forward of measures induced by the inclusion $\kappa : \mathcal{G} \to \mathbb{R}^n$.

Proof. The existence of $\mu^{\mathsf{Hil}}_{H_i(h)} \in \mathcal{DSM}(\mathcal{G})$ follows from the existence of Hilbert decompositions (Oudot & Scoccola, 2024). The equality between measures follows from the uniqueness of $\mu^{\mathsf{Hil}}_{H_i(\kappa \circ h)}$. □

Note that push forwards of discrete measures by any function (such as κ in Lemma D.1) have a very simple expression: all one needs to do is to apply the function to the coordinates of each of the point masses.

Lemma D.2. *Let $f \in \mathsf{Fil}_n(K)$. There exists $p_f, q_f \in \mathbb{N}$ such that, for every $g \in \mathsf{cell}(f)$, the number of positive (resp. negative) point masses of $\mu^{\mathsf{Hil}}_{H_i(g)}$ is p_f (resp. q_f).*

Proof. This follows immediately from Lemma D.1 and the fact that κ is injective, so that no positive-negative pair of masses cancels. □

D.1.2 Proof of semilinearity

Lemma D.2 allows us to index the coordinates of the point masses of $\mu^{\mathsf{Hil}}_{H_i(g)}$ in a convenient way, as follows.

Construction D.3. Let $C \in \mathbb{N}$ be the number of cells of $\mathsf{Fil}_n(K)$, which is finite by Lemma C.1. Choose representatives $f^1, \ldots, f^C \in \mathsf{Fil}_n(K)$ of the cells of $\mathsf{Fil}_n(K)$. Consider the finite set

$$\mathsf{indMass}' := \left\{ \begin{array}{c} (c,s,j) \in \mathbb{N} \times \{+,-\} \times \mathbb{N}, \text{ such that} \\ 1 \le c \le C, \\ 1 \le j \le p_{f^c}, \ \text{ if } s = + \\ 1 \le j \le q_{f^c}, \ \text{ if } s = - \end{array} \right\},$$

and define the disjoint union $\mathsf{indMass} := \{1, \ldots, C\} \amalg \mathsf{indMass}'$.

Given $g \in \mathsf{Fil}_n(K)$, let $1 \le e \le C$ such that $g \in \mathsf{cell}(f^e)$. Using Lemma D.2, consider the unique ordering $x_1, \ldots, x_{p_{f^e}} \in \mathbb{R}^n$ of the positive point masses of $\mu^{\mathsf{Hil}}_{H_i(g)}$ that is compatible with the lexicographic order of $\mathbb{R}^n$. Analogously, let $y_1, \ldots, y_{q_{f^e}} \in \mathbb{R}^n$ be the unique order of the negative point masses of $\mu^{\mathsf{Hil}}_{H_i(g)}$ that is compatible with the lexicographic order of $\mathbb{R}^n$.

Consider the element $\mathsf{sortHil}(g) \in \mathbb{R}^{\mathsf{indMass}}$ defined by

$$\mathsf{sortHil}(c) = \begin{cases} 0 & \text{if } c \ne e \\ 1 & \text{if } c = e, \end{cases}$$

when $c \in \{1, \ldots, C\}$, and by

$$\mathsf{sortHil}(c,s,j) = \begin{cases} 0 & \text{if } c \ne e \\ x_j & \text{if } c = e \text{ and } s = + \\ y_j & \text{if } c = e \text{ and } s = -, \end{cases}$$

when $(c,s,j) \in \mathsf{indMass}'$.

In words, $\mathsf{sortHil}(g)$ contains the information of which cell g belongs to, as well as the coordinates of the point masses of the Hilbert decomposition signed measure of g.

Definition D.4. Let $i \in \mathbb{N}$. Define the ith *sorted Hilbert decomposition* as $\mathsf{sortHil} : \mathsf{Fil}_n(K) \to \mathbb{R}^{\mathsf{indMass}}$, as in Construction D.3.

Proposition D.5. *The sorted Hilbert decomposition is semilinearly determined on grids, and thus semilinear.*

Proof. By Theorem 3.19, it is sufficient to let $f \in \mathsf{Fil}_n(K)$ be arbitrary, and show that the function $\mathsf{push}^{\mathsf{sortHil}}_{\mathsf{ord}_f} : \mathsf{incl}(\mathsf{grid}_f) \to \mathbb{R}^{\mathsf{indMass}}$ is semilinear. This function is in fact affine.

To see this, we let $h = \mathsf{ord}_f$ and use Lemma D.1. This guarantees that each point mass of $\mu^{\mathsf{Hil}}_{H_i(g)}$ lies on top of $\iota_g(x)$ for some element $x \in \mathsf{grid}_f$, and thus varies affinely with respect to $\iota_g \in \mathsf{incl}(\mathsf{grid}_f)$. Moreover, the lexicographic order of the point masses of $\mu^{\mathsf{Hil}}_{H_i(g)}$ does not change as g varies: the order on grid_f induced by pulling back the lexicographic order of $\mathbb{R}^n$ along any inclusion ι_g is the same for all $g \in \mathsf{cell}(f)$. This implies that $\mathsf{sortHil}(g)$ varies affinely with respect to incl_g as long as $g \in \mathsf{cell}(f)$, as required. □

D.2 Semilinearity of the multiparameter persistence landscape

For ease of notation, let us fix a simplicial complex K, a number of parameter $n \geq 1 \in \mathbb{N}$, and a homological dimension $i \in \mathbb{N}$.

We now recall the definition of multiparameter persistence landscape from (Vipond, 2020). Note that (Vipond, 2020) works with general multiparameter persistence modules; we specialize the definition to the case of a homology multiparameter persistence module.

Definition D.6. Let $f \in \mathsf{Fil}_n(K)$. Let $k \geq 1 \in \mathbb{N}$. The kth *multiparameter persistence landscape* of $H_i(f)$ is the function $\lambda^k_f : \mathbb{R}^n \to \mathbb{R}$ defined by

$$\lambda^k_f(z) \coloneqq \sup \left\{ \varepsilon \geq 0 \in \mathbb{R} : \mathsf{rk}_f(z-h, z+h) \geq k \text{ for all } h \in \mathbb{R}^n_{\geq 0} \text{ with } \|h\|_\infty \leq \varepsilon \right\} \in \mathbb{R},$$

where we write $\mathsf{rk}_f = \mathsf{rk}(H_i(f))$, to simplify notation.

By convention, the supremum of an empty set is taken to be zero. Note that this supremum is always finite since the support of $H_i(f) : \mathbb{R}^n \to \mathsf{vec}_{\mathbb{F}}$ is bounded from below.

As a function of $f \in \mathsf{Fil}_n(K)$, the multiparameter landscape takes values in a space of functions $\mathbb{R}^n \to \mathbb{R}$, which is not a finite dimensional vector space. In order to prove a semilinearity result, we consider the evaluation of the landscape at points of $\mathbb{R}^n$, and show that this is semilinear.

Definition D.7. Let $z_0 \in \mathbb{R}^n$. The *evaluated multiparameter persistence landscape* is the descriptor $\mathsf{Fil}_n(K) \to \mathbb{R}$ mapping f to $\lambda^k_f(z_0)$.

In fact, we prove the following stronger result, which, down the line, allows for the optimization of the points over which the landscape is evaluated.

Proposition D.8. *The evaluated multiparameter persistence landscape is linearly determined on grids, and thus semilinear. Moreover, the function* $\mathsf{Fil}_n(K) \times \mathbb{R}^n \to \mathbb{R}$ *mapping* (f, z) *to* $\lambda^k_f(z)$ *is semilinear.*

To simplify notation even more, let us fix $k \geq 1 \in \mathbb{N}$ and let $\lambda_f \coloneqq \lambda^k_f$.

Let $\mathbf{1} = (1, \ldots, 1) \in \mathbb{R}^n$. Multiparameter landscapes can be computed using lines of slope $\mathbf{1}$ as follows (see Lemma 21 of (Vipond, 2020)): $\lambda_f(z) = \sup \{\varepsilon \geq 0 : \mathsf{rk}_f(z - \mathbf{1}\varepsilon, z + \mathbf{1}\varepsilon) \geq k\}$. In particular, we get the following, by the monotonicity of the rank:

$$\lambda_f(z) = \sup \{\min(r, s) : r, s \geq 0 \in \mathbb{R}, \mathsf{rk}_f(z - \mathbf{1}r, z + \mathbf{1}s) \geq k\}. \tag{4}$$

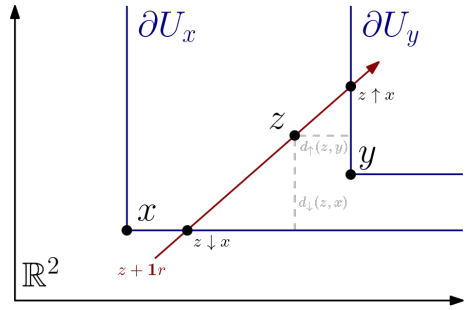


Figure 6. Illustration of the main constructions involved in Proposition D.8.

We now give some useful constructions for the proof of Proposition D.8. Given $x, z \in \mathbb{R}^n$ and $y \in \mathbb{R}^n \cup \{\infty\}$, let

$$d_\downarrow(z,x) := \begin{cases} 0 & \text{if } x \not\leq z \\ \min\limits_{1 \leq j \leq n} |z_j - x_j| & \text{if } x \leq z \end{cases}$$

$$d_\uparrow(z,y) := \begin{cases} \infty & \text{if } y = \infty \\ 0 & \text{if } z \geq y \in \mathbb{R}^n \\ \max\limits_{\substack{1 \leq j \leq n \\ \text{s.t. } z_j \leq y_j}} |z_j - y_j| & \text{if } z \not\geq y \in \mathbb{R}^n \end{cases}$$

$$d_\updownarrow(z,x,y) := \min\left(d_\downarrow(z,x), d_\uparrow(z,y)\right).$$

Given $x, z \in \mathbb{R}^n$, $y \in \mathbb{R}^n \cup \{\infty\}$, with $x \leq z$ and $z \not\geq y$, let

$$z \downarrow x := z - \mathbf{1}d_\downarrow(x,z) \in \mathbb{R}^n$$
$$z \uparrow y := z + \mathbf{1}d_\uparrow(z,x) \in \mathbb{R}^\infty \cup \{\infty\},$$

which have the following property: if $U_w = \{u \in \mathbb{R}^n : u \geq w\} \subseteq \mathbb{R}^n$, then $z \downarrow x$ is the intersection between the line $r \mapsto z + \mathbf{1}r$ and ∂U_x, the boundary of U_x; and $z \uparrow y$ is the intersection between the line $r \mapsto z + \mathbf{1}r$ and ∂U_y. Note that, by definition, we have $\|z - (z \downarrow x)\|_\infty = d_\downarrow(z,x)$ and similarly $\|z - (z \uparrow y)\|_\infty = d_\downarrow(z,y)$.

See Figure 6 for an illustration.

It is standard (and easy to see) that $r \mapsto \mathsf{rk}_f(z - \mathbf{1}r, z + \mathbf{1}r)$ is constant as long as $z - \mathbf{1}r$ and $z + \mathbf{1}r$ do not cross any of the coordinates of the inclusion of the grid of f in $\mathbb{R}^n$; more precisely, as long as $z - \mathbf{1}r$ and $z + \mathbf{1}r$ do not cross the boundary of $U_{\iota_f(a)}$ for some $a \in \mathsf{grid}_f$. From this observation and Equation (4), we get

$$\lambda_f(z) = \max\left\{d_\updownarrow(z, \iota_f(a), \iota_f(b)) : a < b \in \overline{\mathsf{grid}_f}, \mathsf{rk}_{\mathsf{ord}_f}(a,b) \geq k\right\}, \tag{5}$$

where $\overline{\mathsf{grid}_f} = \mathsf{grid}_f \cup \{\infty\}$, the function $\mathsf{rk}_{\mathsf{ord}_f} : \{(a,b) : a < b \in \overline{\mathsf{grid}_f}\} \to \mathbb{Z}$ denotes the rank function of the grid_f filtration $\mathsf{ord}_f : K \to \mathsf{grid}_f$, extended as $\mathsf{rk}_{\mathsf{ord}_f}(a, \infty) := \mathsf{rk}_{\mathsf{ord}_f}(a, \max \mathsf{grid}_f)$, and also extending $\iota_f(\infty) = \infty \in \mathbb{R}^n \cup \{\infty\}$. The extension of grid_f to $\overline{\mathsf{grid}_f}$ is a minor technical point required to handle cases in which a rank does not go below k as its second coordinate goes to ∞, i.e., the case in which $\lambda_f(z) = d_\downarrow(z, \iota_f(a))$ and $\mathsf{rk}_f(z \downarrow \iota_f(a), z + \mathbf{1}r) \geq k$ for all $r \geq 0 \in \mathbb{R}$.

Since grid_f is finite, the supremum of Equation (4) is now a maximum; and again we use the convention that a maximum over an empty set is zero.

Proof of Proposition D.8. Let $f \in \mathsf{Fil}_n(K)$, and let $g \in \mathsf{cell}(f)$. Since $\mathsf{grid}_g = \mathsf{grid}_f$ and $\mathsf{ord}_g = \mathsf{ord}_f$, we have, by Equation (5),

$$\lambda_g(z) = \max\left\{d_\updownarrow(z, \iota_g(a), \iota_g(b)) : a \leq b \in \mathsf{grid}_f, \mathsf{rk}_{\mathsf{ord}_f}(a,b) \geq k\right\},$$

which is a maximum of finitely many semilinear functions that do not depend on g or z, and hence semilinear in both z and ι_g. It follows that the function $\mathsf{Fil}_n(K) \times \mathbb{R}^n \to \mathbb{R}$ mapping (g,z) to $\lambda_g(z)$ is semilinear when restricted to $\mathsf{cell}(f) \times \mathbb{R}^n$ for each $f \in \mathsf{Fil}_n(K)$, and thus semilinear. For the first claim, note that $\mathsf{push}_{\mathsf{ord}_f}(\iota_g) = \lambda_g(z)$, and use Theorem 3.19. $\square$

E Locally Lipschitz objective functions

In this section we prove Proposition E.4, which gives simple conditions under which an objective relying on the Hilbert decomposition signed measure is locally Lipschitz.

In Appendix E.1, we prove a bound that allows us to deduce the stability of signed barcodes as measures descriptors. In Appendix E.2, we use this to give sufficient conditions for a Hilbert decomposition signed measure-based loss function to be locally Lipschitz.

E.1 An algebraic bound

Since this is the only section containing algebraic arguments, and these are encapsulated and only used in the stability result Proposition E.3, we refer the reader to, e.g., (Oudot & Scoccola, 2024; Botnan & Lesnick, 2023) for background.

Lemma E.1. *Fix a finite simplicial complex K and $n \in \mathbb{N}$. There exists a constant B that only depends on K and n such that, for every $f \in \mathsf{Fil}_n(K)$ and $i \in \mathbb{N}$, the sizes of the Betti signed barcode (Oudot & Scoccola, 2024) and of the rank exact decomposition (Botnan et al., 2022b) of the multiparameter persistence module $H_i(f) : \mathbb{R}^n \to \mathsf{vec}_{\mathbb{F}}$ are bounded above by B.*

Proof. We first prove the following.

Claim a. Fix $n \in \mathbb{N}$. Let $P \to Q$ be a map of free n-parameter persistence modules, of ranks $p, q \in \mathbb{N}$, respectively. There exists a constant B' only depending on $n, p, q \in \mathbb{N}$ (and not on P, Q, or the map between them), such that $\mathsf{b}(\mathsf{ker}(P \to Q)) \leq B'$.

Proof of Claim a. This is a consequence of the main result of (Beecher, 2007), in the case where the chosen presentation is not minimal.

Now, for any fixed $f \in \mathsf{Fil}_n(K)$, let $C^f_\bullet$ denote the associated chain complex of n-parameter persistence modules, whose homology is the family of homology persistence modules $H_\bullet(f)$. Note that, for every $k \in \mathbb{N}$, the rank of the free module C^f_k is independent of f. By definition of homology, we have $H_i(f) = \mathsf{coker}(C^f_{i+1} \to Z^f_i)$, where Z^f_i is the kernel of the boundary morphism $Z^f_i = \mathsf{ker}(d^f_i : C^f_i \to C^f_{i-1})$.

To simplify notation, let $\mathsf{b}(M)$ denote the size of the Betti signed barcode of a multiparameter persistence module M, as in (Botnan et al., 2022b).

Claim b. There exists a constant B'', only depending on K and n (and not on f), such that, for all $k \in \mathbb{N}$, $\mathsf{b}(Z^f_k) \leq B''$.

Proof of Claim b. Since C^f_k is non-zero only for finitely many $k \in \mathbb{N}$ (because K is finite), this follows from Claim a and the fact that $Z^f_i = \mathsf{ker}(C^f_i \to C^f_{i-1})$.

We now prove that $\mathsf{b}(H_i(f))$ is bounded above by a constant that only depends on K and n. The bound for the rank exact decomposition of $H_i(f)$ follows from this and Proposition 5.28 of (Botnan et al., 2022b).

Note that $\mathsf{ker}(C_{i+1} \to Z_i) \cong \mathsf{ker}(C_{i+1} \to C_i) = Z_{i+1}$, so we have a short exact sequence

$$0 \to C_{i+1}/Z_{i+1} \to Z_i \to H_i(f) \to 0$$

By Lemma 5.1 of (Botnan et al., 2022b), to bound $\mathsf{b}(H_i(f))$, it is thus enough to bound $\mathsf{b}(C_{i+1}/Z_{i+1})$ and $\mathsf{b}(Z_i)$ by a constant that only depends on K and n. To bound $\mathsf{b}(Z_i)$ we use Claim b.

To bound $\mathsf{b}(C_{i+1}/Z_{i+1})$, we use Lemma 5.1 of (Botnan et al., 2022b), this time with the short exact sequence $0 \to Z_{i+1} \to C_{i+1} \to C_{i+1}/Z_{i+1} \to 0$, and Claim b. This concludes the proof. □

As a direct consequence of Lemma E.1, we get the following.

Corollary E.2. *Fix a finite simplicial complex K and $n \in \mathbb{N}$. There exists a constant B that only depends on K and n such that, for every $f \in \mathsf{Fil}_n(K)$ and $i \in \mathbb{N}$, the number of point masses in both $\mu^{\mathsf{Hil}}_{H_i(f)}$ and $\mu^{\mathsf{rk}}_{H_i(f)}$ is bounded above by B.* □

E.2 Locally Lipschitz signed measure-based objective functions

The next result follows easily from the main results of (Oudot & Scoccola, 2024; Botnan et al., 2022b), and the algebraic bound (Lemma E.1).

Proposition E.3. *Let K be a finite simplicial complex and let $n, i \in \mathbb{N}$. The following functions are Lipschitz with respect to any ℓ^p norm on $\mathsf{Fil}_n(K) \subseteq (\mathbb{R}^n)^K$ and the (partial) optimal transport distance on discrete signed measures:*

$$\mu^{\mathsf{Hil}}_{H_i} : \mathsf{Fil}_n(K) \to \mathcal{DSM}(\mathbb{R}^n)$$
$$\mu^{\mathsf{rk}}_{H_i} : \mathsf{Fil}_n(K) \to \mathcal{DSM}\left(\mathsf{bars}_n, \Delta\right).$$

Proof. From the bottleneck stability of the Betti signed barcode and the rank exact decomposition (Oudot & Scoccola, 2024; Botnan et al., 2022b) we obtain a partial optimal transport (Definition A.2) stability result (sometimes known as 1-Wasserstein stability result in the topological data analysis literature), as long as have a global bound on the number of point masses. This is because the Hilbert decomposition signed measure is obtained from the Betti signed barcode by cancelling equal masses that appear as positive and as negative (Remark 5.3 of (Oudot & Scoccola, 2024)); and the same is true for the rank exact decomposition and the rank decomposition signed barcode. The bound on the number of masses is the content of Lemma E.1. □

Again, for notational clarity, fix $n \geq 1 \in \mathbb{N}$, $i \in \mathbb{N}$, and a simplicial complex K.

Recall that $\mathsf{sortHil}$ denotes the sorted Hilbert decomposition (Definition D.4), which is a convenient representation of the Hilbert decomposition signed measure as a vector of a finite dimensional space.

Proposition E.4. *Assume given a locally Lipschitz function $E' : \mathcal{DSM}(\mathbb{R}^n) \to \mathbb{R}$, and let $E : \mathbb{R}^{\mathsf{indMass}} \to \mathbb{R}$ be any function such that $E' \circ \mu^{\mathsf{Hil}}_{H_i} = E \circ \mathsf{sortHil} : \mathsf{Fil}_n(K) \to \mathbb{R}$. Then, $E \circ \mathsf{sortHil} : \mathsf{Fil}_n(K) \to \mathbb{R}$ is locally Lipschitz.*

Proof. By assumption, $E \circ \mathsf{sortHil} = E' \circ \mu^{\mathsf{Hil}}_{H_i}$, and the right-hand side is a composite of a locally Lipschitz map E', by assumption, and a Lipschitz map $\mu^{\mathsf{Hil}}_{H_i} : \mathsf{Fil}_n(K) \to \mathcal{DSM}(\mathbb{R}^n)$, by Proposition E.3. □

Note that a function E as in Proposition E.4 always exists since all $\mathsf{sortHil}(f)$ is doing is encoding the point masses of the signed measure $\mu^{\mathsf{Hil}}_{H_i}(f)$ in a convenient way. In other words, E is simply an explicit representation of E' when the signed measure is encoded as in Construction D.3.

F Details about experiments

F.1 Details about autoencoder experiment

We use a simple autoencoder architectures with both encoders and decoders made of three layers of 32 neurons, with the first two layers followed by ReLU activation and batch normalization. We then minimize a linear combination of the MSE loss (between initial and reconstructed spaces) and the topological loss (4) (between initial and latent spaces), with weights 1 and 0.1 for the MSE and topological losses respectively, to account for the scale difference between the two losses. Optimization is performed with Adam optimizer, learning rate 0.01, and 1000 epochs.

F.2 Details about graph data example

In order to create node and edge attributes from persistence diagrams[3], the authors in (Horn et al., 2022b) propose to use the natural bijections between persistence diagram points in dimension 0 (resp. dimension 1) with nodes (resp. edges) of the graph[4]. These bijections can be used to permute the persistence diagrams, so that every node (resp. edge) can be associated to its k corresponding persistence diagram points in dimension 0 (resp. dimension 1), and further processed with, e.g., a DeepSet architecture, in order to create a single node vector that is added to the one obtained from graph convolutions. The edge vectors, on the other hand, are pooled so as to create graph-level descriptors.

[3] I.e., one-parameter barcodes, or one-parameter rank decomposition signed measure μ^{rk} (Appendix A.2), in our language.

[4] Note that, in dimension 1, this bijection is only well defined after matching the edges that are not involved in the creation of any graph cycle to an arbitrary persistence diagram point on the diagonal—see Section A.4 in (Horn et al., 2022b)

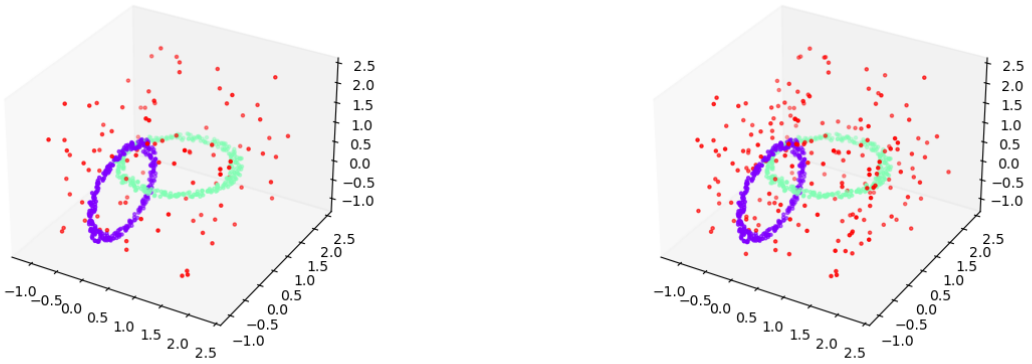


Figure 7. Two point cloud datasets consisting of two interlaced circles with background noise, embedded in $\mathbb{R}^9$, similar to the data used in (Carrière et al., 2021). The difference between the two datasets is the amount of background noise.

Unfortunately there is no such correspondence between nodes or edges and the Hilbert decomposition signed measure in dimension 0 and 1. Hence, for a given node v with associated filtration values $f(v) \in \mathbb{R}^k$ (again, recall that f is obtained from graph convolutions), we create node vectors in $\mathbb{R}^m$ by computing:

$$v \mapsto \int \psi_{v,i} \, \mathsf{d}\mu^{\mathsf{Hil}}_{H_*(f)}, 1 \leq i \leq m, \tag{6}$$

where $\psi_{v,i}$ is a function of the form $\psi_{v,i}(p) = \exp\left(-(p - f(v))^T \Sigma_i^{-1} (p - f(v))\right)$, and the Σ_i's are m learnable SPD matrices in $\mathbb{R}^{k \times k}$. This is thus an instance of Example 3.23. These vectors are then mapped back to $\mathbb{R}^k$ with a fully connected layer in order to add them to the original node attributes, as proposed in (Horn et al., 2022b). We use the same procedure for edges, except that the edge vectors in $\mathbb{R}^m$ are pooled into graph-level descriptors.

For graph neural networks, we use Graph Convolutional Networks (GCN) (Kipf & Welling, 2017), Graph Isomorphism Network (GIN) (Xu et al., 2019), Graph ResNet (He et al., 2016), and Graph DenseNet (Huang et al., 2017), All graph architectures have four layers with 256 neurons, and, whenever topological vectors are used, they are placed after the second layer and computed from $k = 2$ filtrations. In our experience, using a larger k produced comparable results at a higher computational cost. GNNs are trained during 200 epochs with Adam optimizer with learning rate 0.001, and performance is computed over 10 train/test folds.

G Dependency graph and notation table

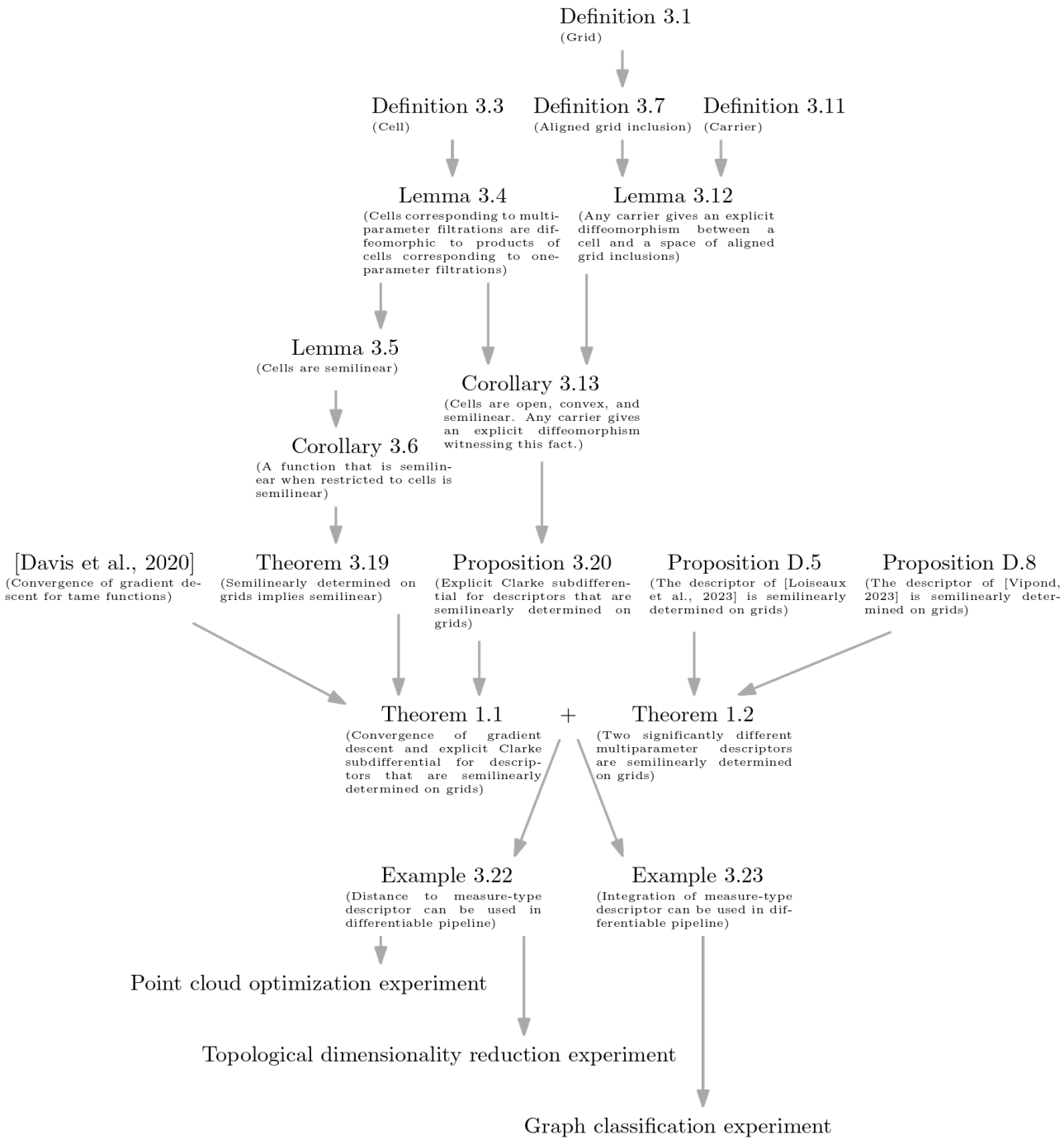


Figure 8. Dependency graph of the main definitions, results, and experiments in the paper.

Symbol	Description
K	A *finite simplicial complex* (page 3).
$(\mathbb{R}^n)^X$	The set of functions $X \to \mathbb{R}^n$ (page 3).
$\mathsf{Fil}_n(K) \subseteq (\mathbb{R}^n)^K$	The *set of n-filtrations* of K (Definition 2.2).
$\mathbb{F}$	A field.
$\Phi : \mathbb{R}^d \to \mathsf{Fil}_n(K)$	A *parameterized family of filtrations*.
$H_i(K)$	The ith *homology* of K with coefficients in $\mathbb{F}$.
$H_i(f)$	The ith *multiparameter persistent homology* of an n-filtration f of K (Definition 2.4).
$\mathsf{Hil}(H_i(f)) : \mathbb{R}^n \to \mathbb{Z}$	The ith *Hilbert function* of $f \in \mathsf{Fil}_n(K)$ (page 3).
$\mathsf{rk}(H_i(f)) : \{(r,s) \in (\mathbb{R}^n)^2 : r \leq s\} \to \mathbb{Z}$	The ith *rank invariant* of $f \in \mathsf{Fil}_n(K)$ (page 4).
M	A metric space.
$\mathcal{DM}(M)$	The set of *discrete measures* on M (page 4).
$\mathcal{DSM}(M)$	The set of *discrete signed measures* on M (page 4).
OT	The *optimal transport distance* on the space of discrete measures (Appendix A.1.1).
$\mu^{\mathsf{Hil}}_{H_i(f)} \in \mathcal{DSM}(\mathbb{R}^n)$	The ith *Hilbert decomposition signed measure* (Definition 2.5).
$S \subset \mathbb{R}^n$	A semilinear set (page 4).
$\partial\mathcal{L}(z)$	The *Clarke subdifferential* of $\mathcal{L}$ at $z \in \mathbb{R}^d$ (page 4).
$[m]$	The *linear order* $\{0 < 1 < \cdots < m-1\}$ for $m \in \mathbb{N}$.
$\mathcal{G} = \mathcal{G}_1 \times \cdots \times \mathcal{G}_n$	A *grid* (Definition 3.1).
$\mathsf{grid}_f = [m_1] \times \cdots \times [m_n]$	The *grid* induced by a filtration f (Construction 3.2).
$\mathsf{ord}_f : K \to \mathsf{grid}_f$	The *unique linear preorder* induced by a filtration f (Construction 3.2).
$\mathsf{cell}(f) \subseteq \mathsf{Fil}_n(K)$	The *cell* of $f \in \mathsf{Fil}_n(K)$ (Definition 3.3).
$\mathsf{incl}([m])$	The set of *aligned grid inclusions* of $[m]$ into $\mathbb{R}$ (Construction 3.8).
$\mathsf{incl}(\mathcal{G})$	The set of aligned grid inclusions of a grid $\mathcal{G}$ into $\mathbb{R}^{[m_1]} \times \cdots \times \mathbb{R}^{[m_n]}$ (Construction 3.8).
$C : \mathsf{grid}_f \to K$	A *carrier* for $f \in \mathsf{Fil}_1(K)$ (Definition 3.11).
$\{C_i : \mathsf{grid}_{f_i} \to K\}_{1 \leq i \leq n}$	A *carrier* for a multifiltration $f \in \mathsf{Fil}_n(K)$ (Definition 3.11).
$\mathsf{itsIncl} : \mathsf{cell}(f) \to \mathsf{incl}(\mathsf{grid}_f)$	Function mapping a filtration in the cell of f to its corresponding grid inclusion (Definition 3.10).
$\mathsf{Fil}_{\mathcal{G}}(K)$	The *set of $\mathcal{G}$-filtrations* on K (page 6).
$\mathrm{VR}(X)$	The *Vietoris–Rips filtration* of a finite point cloud $X \subset \mathbb{R}^n$.

Table 2. Notation Table.