MULTI-PARAMETER TOPOLOGICAL VAEs FOR CONDITIONING IMAGE GENERATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Topological Data Analysis (TDA) is the field of data science whose aim is to detect and encode topological patterns in data sets. As such, it has become common in machine learning to complement traditional models with TDA-based features and losses to enforce geometrical constraints into the models behaviour. However, while several methods can ensure that topology can be preserved at a *global* level (eg, total number of components or cycles), constraining topology *locally* (eg, imposing the presence of topological features at specific areas in space) has been relatively unexplored, despite its usefulness in several application domains.

In this article, we propose to fill this gap with *multi-parameter persistent homology*. More precisely, we leverage a recently proposed descriptor, called the *Multi-parameter Module Approximation* (MMA), to design new losses for generative models. Indeed, MMA descriptors are easily interpretable and are thus more amenable to the construction of losses than other standard descriptors. On the theoretical side, we prove that MMA descriptors are locally Lipschitz and definable (under suitable assumptions), so that stochastic subgradient descent converges, which allows us to use them for steering neural net behaviours at training time. On the practical side, we showcase the efficiency of our localized topological constraints for mask and image generation with variational autoencoders, by providing superior results over both single-parameter and non-topological baselines.

1 INTRODUCTION

Topological Data Analysis (TDA) is a field of data science whose aim is to detect and encode the *topological features* of datasets Chazal & Michel (2021); Dey & Wang (2022); this includes connected components (which are features in degree 0), loops (degree 1), enclosed spheres (degree 2), and other higher dimensional cycles. TDA is based on algebraic topology, and more precisely on persistent homology (PH); thanks to these mathematical roots, most TDA descriptors enjoy robustness properties, often called *stability theorems*. Moreover, TDA descriptors are highly versatile: they are computed from a continuous *filtration* function f, whose choice is most of the time left to the user, (although automatic tuning is possible) and that acts as the lens with which one observes the data[1]. As such, TDA descriptors have been used extensively across many fields of data science using appropriate filtration functions, including applications in bioinformatics Kuchroo et al. (2023); Multari et al. (2026); Yang et al. (2025); Kandror et al. (2026), computer graphics Skraba et al. (2010); Carrière et al. (2015); Poulenard et al. (2018), cosmology Biagetti et al. (2022); Yip et al. (2024; 2025), and time series Perea & Harer (2015); Zeng et al. (2021); Chazal et al. (2024), to name a few.

The standard TDA descriptor is *persistent homology* (PH), which filters a dataset X through the sublevel sets of a function $f\colon X \to \mathbb{R}$ and records the value at which each topological feature appears and the value at which it disappears. A recent important progress in the TDA community is the ability to deal with more than one filtration at the time, or, equivalently, with multi-parameter functions $f\colon X \to \mathbb{R}^n$, $n > 1$, allowing to detect and encode topological features that are shared across several lenses, a field called *multi-parameter PH* Carlsson & Zomorodian (2009); Botnan & Lesnick (2023). Typical examples include the joint processing of scale and density, in order to

[1]Indeed, topological features are detected if their *sizes*, that is, the diameters of their ranges in the function image, are large enough. On the other hand, a "horizontal" cycle on which f is constant is harder to capture.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

detect topological features that are present at several scales even in the face of outlier points (whereas single-parameter PH would be highly sensitive to outlier noise). Even though the theoretical aspects of multi-parameter PH are substantially more involved, tremendous progress has been made in the recent years towards the goal of providing efficient guarantees Vipond (2020); Mukherjee et al. (2026); Loiseaux et al. (2023b; 2025; 2023a); Kerber & Russold (2025).

In parallel, a large effort has also been devoted to the study of the differentiability properties of PH, i.e., the possibility to properly define *gradients* for PH-based descriptors Carrière et al. (2021); Leygonie et al. (2023); Nigmetov & Morozov (2024), in order to incorporate them into the auto-differentiation framework of neural networks, whether it is for training models with topological data signatures or to steer model behaviors by imposing topological constraints to losses at training time. See Carrière et al. (2026) for a recent survey. A natural application, which we further develop in this article, is the case of *topologically constrained mask generation*: in Gupta et al. (2025), diffusion models are trained to generate masks whose topology is imposed beforehand, therefore ensuring that the produced masks have the right numbers of components and cycles.

Challenges in TDA. In this article, we focus on two main bottlenecks. The first one is theoretical, and relates to multi-parameter TDA descriptors, and the second one is practical, and focuses on mask generation with topological constraints.

Developing differentiable versions of multi-parameter PH descriptors has recently been studied Mukherjee et al. (2026); Scoccola et al. (2024); Korkmaz et al. (2025). However, compared to the single-parameter setting, these efforts are mostly preliminary. Of particular interest to us are descriptors computed with ***Multi-parameter Module Approximations*** **(**MMA**) Loiseaux et al. (2025), whose differentiability properties are currently missing in the literature**. This lack is critical in the context of topological constraints for neural networks, as, compared to MMA descriptors, the high complexity and level of abstraction of the other descriptors, such as the signed measure Scoccola et al. (2024) and the generalized rank invariant Mukherjee et al. (2026), makes it difficult to use them for designing new multi-parameter topological constraints in losses. On the other hand, MMA descriptors enjoy higher interpretability: they are comprised of *summands*, i.e., topological feature lifetimes, exactly like their single-parameter equivalents, the persistence diagrams. This makes them appropriate for easily deriving new topological terms in losses.

The mask generation models developed in Gupta et al. (2025) are naturally limited by their use of single-parameter PH. **They can make sure the *total* number of topological features is preserved, but they cannot impose *where* these features should be placed in the masks**. This can only be constrained by using a second filtration when building topological descriptors, that encodes where features should be located, and by designing an appropriate loss to minimize in order to satisfy these *local* topological constraints.

Contributions. In this article, our contributions are threefold:

1. MMA **optimization.** We first prove that MMA descriptors are locally Lipschitz on an open dense set of filtrations of full measure, where MMA-based losses therefore admit Clarke subgradients, providing us with a theoretical basis for their use in gradient-based topological constraining of neural networks. See our main result Theorem 12.
2. **Location filtration and loss.** Second, we introduce a new filtration to be used in addition to the image filtration, that characterizes where topological features should be placed, and incorporate it to a conditional VAE (CVAE) with an appropriate MMA-based loss. We call the resulting model *Topo-CVAE*, see Figure 1 and Section 4.1. The two filtrations enter a single bifiltration per object, whose MMA summands carry the topology of that object at its prescribed location, so that the loss imposes a separate topological requirement at every position.
3. **Improved conditional generation.** Finally, we train Topo-CVAE for mask generation, that are constrained on the presence *and* locations of topological features, and show that our resulting model outperforms both single-parameter PH-based and non-topology-based baselines. See Section 4.2.

Related work. Combining PH with neural networks has mostly been done with single-parameter PH in the literature, and applications roughly fall into three categories: (a) improving model per-

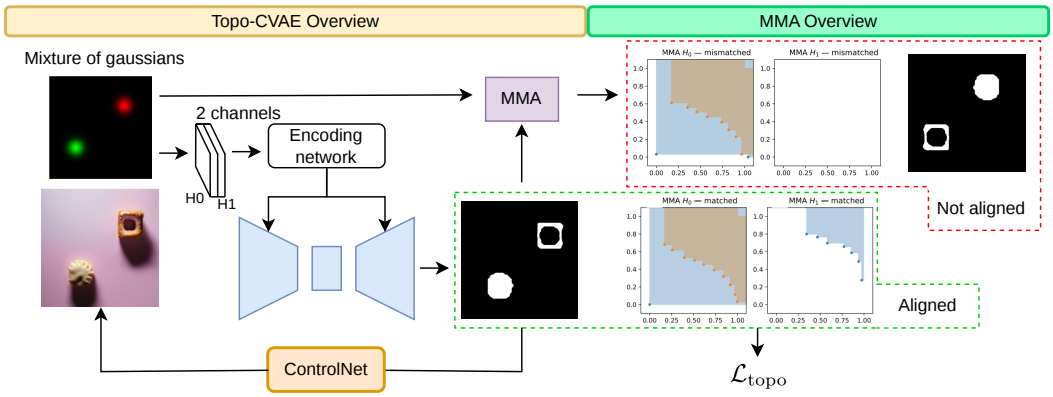


Figure 1: **Topo-CVAE overview (left):** Topo-CVAE generates binary masks in which each object has a prescribed location and topology (filled or with a hole). The conditioning map is a mixture of Gaussians, one per object, split into an H_0 channel for filled objects (green) and an H_1 channel for objects that should contain a hole (red). MMA is applied to the generated mask and the conditioning map, and its summands define the topological loss $\mathcal{L}_{\text{topo}}$; the mask can then be passed to ControlNet Zhang et al. (2023) to generate the final image. MMA **overview (right):** we compute a 2-parameter MMA descriptor from the image intensity and the conditioning map, in homology degrees 0 and 1, for two masks built from the same filled disk and the same hollow square. When each object sits at the location it is meant to occupy (*Aligned*, green box), a degree-1 summand appears, which $\mathcal{L}_{\text{topo}}$ rewards. When the two objects are swapped (*Not aligned*, red box), the degree-1 descriptor is empty: the hole in the mask is not where the conditioning map prescribes it.

formances by feeding neural nets with topological descriptors Carrière et al. (2020); Hofer et al. (2020); Horn et al. (2022); Zeng et al. (2024; 2021); (b) monitoring neural networks with topological signatures encoding the geometric patterns in network weights or loss landscapes Rieck et al. (2019); Lacombe et al. (2021); Barannikov et al. (2025), and (most relevant to us) (c) steering the model behavior at training time with topological losses, which give latent spaces the topology of the input data in *dimensionality reduction* Moor et al. (2020); Vandaele et al. (2022); Carrière et al. (2024), control the topological complexity of classifiers for *generalization* Chen et al. (2019); Hacquard et al. (2022), and prescribe the topology of the samples in *data generation* Wang et al. (2020); Clough et al. (2022); Gupta et al. (2025). See Ballester et al. (2026) for a survey.

Comparatively, combining multi-parameter PH with neural networks has been less studied, and mostly amounts to learning the best filtrations to use in order to produce better topological descriptors Scoccola et al. (2024); Carrière et al. (2025); Mukherjee et al. (2026). Topological model steering has never been done with multi-parameter PH, as the generalized rank invariant and signed barcodes characterize PH too abstractly to yield natural losses. MMA descriptors Loiseaux et al. (2025) are a good candidate alternative, thanks to their higher interpretability, but to the best of our knowledge the literature addresses neither their differentiability properties nor their combinations with neural networks.

Concerning the data generation application in (c) above, most topological models manage to constrain how many features a sample contains, but never *where* they are. In segmentation, Oner et al. (2023) add location information to single-parameter PH losses through a filtration that fuses intensity thresholding with height functions. Conversely, spatial mechanisms such as AIR (Eslami et al., 2016), Spatial Broadcast Decoders (Watters et al., 2019) and CoordConv (Liu et al., 2018) are able to place objects, grounded diffusion models such as GLIGEN (Li et al., 2023) place them at prescribed bounding boxes, and CountGen (Binyamin et al., 2025) generates a prescribed number of them, but none of these controls the data topology.

2 Background

This section introduces multi-parameter PH, the Multi-parameter Module Approximation (MMA) from Loiseaux et al. (2025), and some notions from optimization theory used in Section 3. Through-

out, $\mathcal{K}$ is a finite simplicial complex (e.g., an image whose pixels are the 0-simplices of $\mathcal{K}$, and whose pixel adjacency prescribes the higher-dimensional simplices) and $\Bbbk$ a fixed field. For $n \in \mathbb{N}$ we write $[n] = \{1, \dots, n\}$ and $\bar{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$; the space $\mathbb{R}^n$ carries the product partial order, $a \leq_n b$ if and only if $a_i \leq b_i$ for every $i \in [n]$. We first introduce *persistence modules*.

Persistence modules. An *n-dimensional persistence module* $\mathbb{M}$ is a family of $\Bbbk$-vector spaces $\{M_a\}_{a \in \mathbb{R}^n}$ together with *transition maps* $\varphi_{\mathbb{M}}(a, b) \colon M_a \to M_b$ for $a \leq_n b$, satisfying $\varphi_{\mathbb{M}}(b, c) \circ \varphi_{\mathbb{M}}(a, b) = \varphi_{\mathbb{M}}(a, c)$ and $\varphi_{\mathbb{M}}(a, a) = \mathrm{id}$; it is *pointwise finite dimensional* (p.f.d.) if $\dim M_a < \infty$ for every $a \in \mathbb{R}^n$. Two modules are *ε-interleaved* if each admits a morphism into the ε-shift of the other, in a way compatible with the transition maps, and the *interleaving distance* d_I is the infimum of the ε for which this holds (Lesnick, 2015). When $n = 1$, every p.f.d. module decomposes essentially uniquely into *interval modules* (Crawley-Boevey, 2015): the resulting multiset of intervals, the *persistence diagram*, also called *barcode*, is a complete discrete invariant, and the Isometry Theorem gives $d_I = d_B$, the bottleneck distance between barcodes (Lesnick, 2015, Thm. 3.4). In practice, such modules arise from applying the homology functor to *filtrations*, as described below.

Multi-parameter PH. An *n-dimensional filtration* of $\mathcal{K}$ is a monotonic map $f \colon \mathcal{K} \to \mathbb{R}^n$, that is, $f(\sigma) \leq_n f(\tau)$ whenever $\sigma \subseteq \tau$. We write $\mathrm{Fil}_n(\mathcal{K}) \subseteq \mathbb{R}^{n|\mathcal{K}|}$ for the space of such filtrations, equipped with the supremum norm, and $\mathbb{M}(f)$ for the module obtained by applying the homology functor $H_\bullet(\,\cdot\,; \Bbbk)$ to the sublevel sets $S_x(\mathcal{K}) \coloneqq \{\sigma \in \mathcal{K} \mid f(\sigma) \leq_n x\}$. For $n \geq 2$ no analogue of the barcode is available: the isomorphism classification of finitely generated modules involves continuous moduli (Carlsson & Zomorodian, 2009), which rules out any complete discrete invariant and forces the use of incomplete descriptors. The module itself is nonetheless stable, as $d_I(\mathbb{M}(f), \mathbb{M}(g)) \leq \|f - g\|_\infty$ (Lesnick, 2015, Thm. 5.3).

Interval modules. The *support* of $\mathbb{M}$ is $\mathrm{supp}(\mathbb{M}) \coloneqq \{x \in \mathbb{R}^n \mid M_x \neq 0\}$, and the *direct sum* $\mathbb{M} \oplus \mathbb{M}'$ is defined pointwise. A non-empty $I \subseteq \mathbb{R}^n$ is an *interval*[2] if $a, c \in I$ and $a \leq_n b \leq_n c$ imply $b \in I$; the associated *interval module* $\mathbb{I}$ has $\mathbb{I}_a = \Bbbk$ for $a \in I$ and 0 otherwise, with identity transition maps inside I. We call $\mathbb{I}$ *discretely presented* when I is a locally finite union of rectangles, that is, when there are locally finite families of *birth corners* B and *death corners* D in $\bar{\mathbb{R}}^n$ with

$$I = \bigcup_{b \in B} \bigcup_{d \in D} R_{b,d}, \qquad R_{b,d} \coloneqq \{x \in \bar{\mathbb{R}}^n \mid b \leq_n x <_n d\}, \tag{1}$$

where $R_{b,d} = \emptyset$ whenever $b \not\leq_n d$. All interval modules below are discretely presented, and we write $\mathcal{I}$ for the class of finite direct sums of such modules, each recorded together with its summands and their lists of birth and death corners.

Fibered barcode. A *diagonal line* is a line of the form $l_b \coloneqq b + t\mathbf{1}$, $t \in \mathbb{R}$, with basepoint $b \in \mathbb{R}^{n-1} \times \{0\}$; throughout, L denotes a finite set of equally spaced diagonal lines with spacing $\delta > 0$ (Appendix A). Restricting $\mathbb{M}(f)$ to $l \in L$ yields the 1-parameter module induced by the filtration $f^l \colon \mathcal{K} \to \mathbb{R}$, $f^l(\sigma) \coloneqq \max_{i \in [n]}(f(\sigma)_i - b_i)$, which in turn induces the total preorder $\sigma \leq_{f,l} \tau$ if and only if $f^l(\sigma) \leq f^l(\tau)$. Each restriction is p.f.d. and therefore has a barcode $\mathcal{B}(\mathbb{M}(f)|_l)$; the family $\{\mathcal{B}(\mathbb{M}(f)|_l)\}_{l \in L}$ is the *L-fibered barcode* of $\mathbb{M}(f)$ (Lesnick & Wright, 2015; Landi, 2018).

Multi-parameter Module Approximation. Given a persistence module $\mathbb{M}$ and $l, l' \in L$, a *matching function* $\rho \colon \mathcal{B}(\mathbb{M}|_l) \to \mathcal{B}(\mathbb{M}|_{l'}) \cup \{\emptyset\}$ is a map whose restriction to $\rho^{-1}(\mathcal{B}(\mathbb{M}|_{l'}))$ is injective. By abuse of notation we also write ρ for a family of such maps, one for each pair of lines of L; which pair being clear from the context. Given $f \in \mathrm{Fil}_n(\mathcal{K})$, the *Multi-parameter Module Approximation* (Loiseaux et al., 2025, Alg. 1) is the finite direct sum of discretely presented interval modules

$$\mathrm{MMA}_L(f, \rho) \coloneqq \bigoplus_{i \in [m]} \mathbb{I}_i^f \in \mathcal{I}, \tag{2}$$

obtained as follows: ρ groups the bars of the L-fibered barcode into families, one per summand; the endpoints of the bars in family i give its birth and death corners B_i^f, D_i^f, and its support is given by Equation (1). Figure 2 illustrates the construction. We drop ρ from the notation when it is the matching produced by the vineyard algorithm (Cohen-Steiner et al., 2006). The best available guarantee is up to the grid spacing: for suitable compatible matchings ρ_f, ρ_g one has

[2] This is weaker than Loiseaux et al. (2025, Def. 6), which also requires I to be connected; see Appendix A.

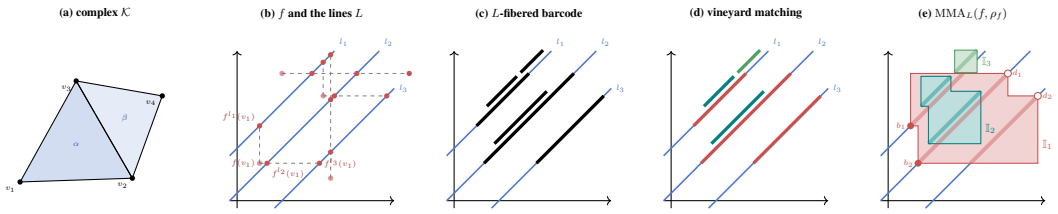


Figure 2: Construction of MMA_L. (a) A filtered simplicial complex. (b) The filtration values in $\mathbb{R}^2$, the set L of diagonal lines, and the induced values $f^l(\sigma)$ obtained by projecting onto each line, shown here for the simplex v_1. (c) The resulting L-fibered barcode. (d) The vineyard matching groups bars across consecutive lines into families, one colour per family. (e) Each family gives one interval summand; for $\mathbb{I}_1$ the birth corners b_1, b_2 (filled) and death corners d_1, d_2 (hollow) are marked, and the support is the union of the rectangles they span.

$d_I(\mathrm{MMA}_L(f, \rho_f), \mathrm{MMA}_L(g, \rho_g)) \leq \|f - g\|_\infty + \delta$ (Loiseaux et al., 2025, Prop. 4). Throughout, we use the simplified corner computation of Loiseaux et al. (2025, Rmk. 8), which takes bar endpoints directly as corners, so that the number of corners depends only on the simplex pairing and not on the filtration values.

Optimization of tame functions. An *o-minimal structure* is a sequence of Boolean algebras $\mathcal{O}_n$ of subsets of $\mathbb{R}^n$, stable under products and coordinate projections, containing the zero sets of all polynomials, and whose elements in $\mathcal{O}_1$ are exactly the finite unions of points and intervals; its sets, and the maps whose graph is one, are called *definable* (Appendix B). A set is *semiaffine* if it is a finite union of sets cut out by finitely many affine equalities and strict inequalities; semiaffine sets are definable in any o-minimal structure, and any piecewise affine map is semiaffine (Van den Dries, 1998). Definability is what makes stochastic subgradient descent converge: for $F\colon \mathbb{R}^d \to \mathbb{R}$ locally Lipschitz and definable, under the standard assumptions on the learning rate and the noise (Appendix B), almost surely every limit point of the iterates is critical for F and the values $F(x_n)$ converge (Davis et al., 2020, Cor. 5.9).

3 A DIFFERENTIABLE FRAMEWORK FOR MMA

As explained in Section 2, MMA descriptors are quite natural, as they are comprised of summands whose supports represent approximate lifetimes of topological features. However, optimising through MMA_L requires a locally Lipschitz and definable objective, and neither is available: definability is not defined on $\mathcal{I}$, and the enforced stability of Proposition 4 of Loiseaux et al. (2025) retains an additive δ that survives the limit $\|f - g\|_\infty \to 0$. Blocks (Section 3.1) resolve the first, by exhibiting the descriptor as a piecewise affine map into a Euclidean space on each of finitely many semiaffine charts; a genericity assumption (Section 3.2) resolves the second. Together they place the pipeline within the scope of Proposition 18, for a class of vectorisations that we characterise in Section 3.3. All proofs of this section, together with auxiliary results, are deferred to Appendix C.

3.1 BLOCKS

The first ingredient is a single combinatorial construction: a partition of $\mathrm{Fil}_n(\mathcal{K})$ into families of filtrations on which the MMA descriptor behaves uniformly. Throughout this section we fix L and $\mathcal{K}$. Orders with ties are completed as in Loiseaux et al. (2025, Eq. (22)).

Definition 1. Given $f \in \mathrm{Fil}_n(\mathcal{K})$, the **block** of f is defined as

$$\mathrm{block}_L(f) := \bigcap_{l \in L} \{g \in \mathrm{Fil}_n(\mathcal{K}) \mid \leq_{f,l} \equiv \leq_{g,l}\}.$$

Proposition 2. *The blocks are finite in number and form a partition of* $\mathrm{Fil}_n(\mathcal{K})$.

Proposition 3. $\mathrm{block}_L(f)$ *is semiaffine for every* $f \in \mathrm{Fil}_n(\mathcal{K})$.

To turn blocks into charts we record the combinatorial type of the output. A **configuration** is a tuple $\mathcal{C} = (m, p_1, q_1, \ldots, p_m, q_m)$ recording the number m of interval summands and the numbers p_i, q_i of birth and death corners of the i-th one. The corners of such a family are collected

in the **configuration space** $\mathcal{S}_{\mathcal{C}} \coloneqq \prod_{i=1}^{m}(\mathbb{R}^{np_i} \times \mathbb{R}^{nq_i})$, and the embedding $e_{\mathcal{C}} \colon \mathcal{S}_{\mathcal{C}} \to \mathcal{I}$ sends $((B_1, D_1), \dots, (B_m, D_m))$ to the direct sum of the intervals they span through Equation (1). Corners are allowed to take values in $\bar{\mathbb{R}}$; which coordinates are infinite is determined by the simplex pairing, hence constant on each block, so we identify $\mathcal{S}_{\mathcal{C}}$ with the Euclidean space of the finite ones. We now show that the configuration is constant on each block, which is what makes the partition useful.

Lemma 4. *The configuration of* $\mathrm{MMA}_L(f)$ *is constant on* $\mathrm{block}_L(f)$.

This allows us to factor MMA_L through the Euclidean space $\mathcal{S}_{\mathcal{C}}$ on each block.

Proposition 5. *Let* $f \in \mathrm{Fil}_n(\mathcal{K})$ *and let* $\mathcal{C}$ *be the configuration of* $\mathrm{MMA}_L(f)$. *There exists a semiaffine, locally Lipschitz map* $\kappa \colon \mathrm{block}_L(f) \to \mathcal{S}_{\mathcal{C}}$ *with* $\mathrm{MMA}_L = e_{\mathcal{C}} \circ \kappa$ *on* $\mathrm{block}_L(f)$.

3.2 Local Lipschitz continuity of MMA

From here on we fix $n = 2$, we take ρ_h to be the matching produced by the vineyard algorithm, which is compatible under Assumption 9 by Remark 32 and Loiseaux et al. (2025, Prop. 17), and we work in the setting of Remark 8 recalled in Section 2. Since $n = 2$, the lines of L can be ordered by their basepoints, $L = (l_1, \dots, l_N)$. No Lipschitz estimate for MMA_L is available in the literature: the closest statement, Loiseaux et al. (2025, Prop. 4), carries an additive δ, and Section 4.1 of the same paper documents an instability of MMA for arbitrary compatible matchings on modules that are close in d_I. The setting below is a different one, in which such an estimate does hold. The results of this section rest on three assumptions, which we now state: Assumption 6 underlies all of them, and Assumptions 8 and 9 enter Theorem 11.

Assumption 6. Let $\Omega \subseteq \mathrm{Fil}_n(\mathcal{K})$ be the set of filtrations h such that $h^l(\varsigma_1) \neq h^l(\varsigma_2)$ for every $l \in L$ and every pair of distinct simplices $\varsigma_1, \varsigma_2 \in \mathcal{K}$. We assume $f, g \in \Omega$.

The assumption bears on f and g alone, and not on the filtrations between them. It is mild: Ω is dense in $\mathrm{Fil}_n(\mathcal{K})$ and its complement has Lebesgue measure zero, being contained in a finite union of hyperplanes. This mirrors the one-parameter setting, where a bar that shrinks to zero length does not disappear but reaches the diagonal: the number of points of a persistence diagram is determined by $\mathcal{K}$ alone, and the persistence map is a permutation of coordinates, constant on each piece of a semialgebraic partition of the space of filtrations and Lipschitz there (Carrière et al., 2021, §2.2, Prop. 3.2). The convergence guarantees of topological optimisation rest on exactly this structure.

Definition 7. For $l \in L$ and distinct $\varsigma_1, \varsigma_2 \in \mathcal{K}$, the **boundary** $W^l_{\varsigma_1\varsigma_2}$ is the set of filtrations $h \in \mathrm{Fil}_n(\mathcal{K})$ with $h^l(\varsigma_1) = h^l(\varsigma_2)$.

By definition, Ω is the complement in $\mathrm{Fil}_n(\mathcal{K})$ of the union of all boundaries.

Assumption 8. The segment $[f, g]$ meets every boundary in finitely many points, no point of $[f, g]$ lies on two distinct boundaries, and $[f, g]$ meets no boundary $W^{l_1}_{\varsigma_1\varsigma_2}$ of the first line of L.

Assumption 9. For every $h \in [f, g] \cap \Omega$ and every pair l, l' of consecutive lines of L, the orders $\leq_{h,l}$ and $\leq_{h,l'}$ coincide or differ by one adjacent transposition.

Assumption 9 says that at most one adjacent transposition of the simplex order occurs between two consecutive lines of L. For a fixed L it can fail on an open set of filtrations. Appendix D shows that it holds on a neighbourhood of every filtration whose exchange points, including those of the pairs of simplices that small perturbations make incomparable, lie more than δ apart, hence near almost every filtration once δ is small enough, and that the vineyard matching remains compatible under it.

Theorem 10. *Let* $f \in \Omega$ *and set*

$$\varepsilon_f \coloneqq \tfrac{1}{2} \min\big\{|f^l(\varsigma_1) - f^l(\varsigma_2)| \;:\; l \in L,\ \varsigma_1 \neq \varsigma_2 \in \mathcal{K}\big\} > 0.$$

Then $B(f, \varepsilon_f) \cap \mathrm{Fil}_n(\mathcal{K}) \subseteq \mathrm{block}_L(f)$, *and consequently*

$$d_I\big(\mathrm{MMA}_L(g), \mathrm{MMA}_L(g')\big) \leq \|g - g'\|_\infty \qquad \text{for all } g, g' \in B(f, \varepsilon_f) \cap \mathrm{Fil}_n(\mathcal{K}).$$

In particular, MMA_L *is locally* 1*-Lipschitz from* $(\mathrm{Fil}_n(\mathcal{K}), \|\cdot\|_\infty)$ *to* $(\mathcal{I}, d_I)$ *at every point of* Ω.

Theorem 10 covers an open dense set of full measure. As the filtration moves, the orders of the simplices on the lines of L change, and with them the configuration of MMA_L. Theorem 11 controls these changes whenever the segment joining two filtrations of Ω crosses the block boundaries as prescribed by Assumptions 8 and 9.

Theorem 11. *Let $f, g \in \Omega$ be such that the segment $[f, g]$ satisfies Assumptions 8 and 9. Then*

$$d_I\big(\mathrm{MMA}_L(f), \mathrm{MMA}_L(g)\big) \leq \|f - g\|_\infty.$$

3.3 From MMA_L to a differentiable loss

To feed the descriptor to a learning pipeline one maps it to a Euclidean space. Given $D \in \mathbb{N}$, a **vectorisation** is a map $\mathrm{vec}\colon \mathcal{I} \to \mathbb{R}^D$. Throughout this section we fix an o-minimal structure $\mathcal{O}$, and definability is understood with respect to it.

For $s, s' \in \mathcal{S}_\mathcal{C}$, the pseudometric $d_\mathcal{C}(s, s')$ is the minimum, over the bijections between their summands that preserve the numbers of corners and the infinite coordinates, of the largest $\|\cdot\|_\infty$-distance between corresponding finite corners.

Theorem 12. *Let $\mathcal{L} \coloneqq E \circ \mathrm{vec} \circ \mathrm{MMA}_L \circ \Phi$, with $\Phi\colon \mathbb{R}^d \to \mathrm{Fil}_n(\mathcal{K})$, $\mathrm{vec}\colon \mathcal{I} \to \mathbb{R}^D$ and $E\colon \mathbb{R}^D \to \mathbb{R}$. Assume that (1) Φ and E are definable and locally Lipschitz and that, for every configuration $\mathcal{C}$, (2) $\mathrm{vec} \circ e_\mathcal{C}$ is definable and (3) $E \circ \mathrm{vec} \circ e_\mathcal{C}$ is locally Lipschitz on $(\mathcal{S}_\mathcal{C}, d_\mathcal{C})$. Then $\mathcal{L}$ is definable and, for every $x \in \mathbb{R}^d$ with $\Phi(x) \in \Omega$: (i) $\mathcal{L}$ is locally Lipschitz at x; (ii) there exist $r > 0$ and $\Lambda' \geq 0$ such that*

$$|\mathcal{L}(x) - \mathcal{L}(x')| \leq \Lambda' \, \|\Phi(x) - \Phi(x')\|_\infty$$

for every $x' \in \mathbb{R}^d$ with $\Phi(x') \in \Omega$, $\|\Phi(x) - \Phi(x')\|_\infty < r$, and $[\Phi(x), \Phi(x')]$ satisfying Assumptions 8 and 9.

Condition (3) asks for regularity in the corners of the summands, grouped by the vineyard matching. It holds whenever $E \circ \mathrm{vec}$ is locally Lipschitz for d_I, since $d_I \leq d_\mathcal{C}$ by Lemma 24, and it also holds for losses on individual summands such as Equation (5) (Proposition 31).

Corollary 13. *Let $\Phi\colon \mathbb{R}^d \to \Omega$ be definable and locally Lipschitz, let $E\colon \mathbb{R}^D \to \mathbb{R}$ and $\mathrm{vec}\colon \mathcal{I} \to \mathbb{R}^D$ be such that $E \circ \mathrm{vec} \circ e_\mathcal{C}$ is definable and locally Lipschitz on $\mathcal{S}_\mathcal{C}$ for every configuration $\mathcal{C}$, and set $\mathcal{L} \coloneqq E \circ \mathrm{vec} \circ \mathrm{MMA}_L \circ \Phi$. Then $\Phi(\mathbb{R}^d)$ lies in a single block, $\mathcal{L}$ is definable and locally Lipschitz on $\mathbb{R}^d$, and, under Assumption 17, almost surely every limit point of the iterates of the stochastic subgradient method applied to $\mathcal{L}$ is critical for $\mathcal{L}$, and the values $\mathcal{L}(x_n)$ converge.*

Within a block, the vineyard matching fixes how the corners are grouped into summands, and Corollary 13 reads the loss in these grouped coordinates. Across a boundary described by Assumptions 8 and 9, the matching exchanges two corners of the same kind on one line of L; the proof of Theorem 11 bounds the resulting change of the corners by the change of the filtration, and condition (3) of Theorem 12 carries this bound to $\mathcal{L}$ through item (ii).

4 Experiments

In this section, following Gupta et al. (2025), we use our differentiable version of MMA descriptors to train conditional VAEs for mask generation with prescribed topology, using an appropriate MMA-based loss to steer the model at training time. Note that our framework differs slightly from the one of Gupta et al. (2025): our numerical experiments illustrate how biased or inefficient models can be fixed with MMA-based losses without increasing the model sizes. Indeed, as deep spatially conditioned U-Net-based diffusion models Gupta et al. (2025) can already produce masks with the requested topology without using topological losses, we rather focus on models that underperform and show that adding our MMA-based loss improves model adherence to local topological requirements, in a more efficient way than both 1-parameter PH-based and topology-agnostic models.

4.1 Model

Conditional variational autoencoder. We generate binary masks using a conditional variational autoencoder (CVAE) Sohn et al. (2015) with a SPADE-conditioned decoder Park et al. (2019). Its two-channel input c encodes object positions through mixtures of Gaussians centered at the prescribed locations: channel 0 represents filled objects (H_0), while channel 1 represents hollow objects with a hole (H_1). This conditioning map specifies locations and topology without directly providing object boundaries. During training, we sample $z \sim q_\phi(z \mid x, c)$ and decode foreground probabilities $\hat{x} = D_\theta(z, c)$; during inference, $z \sim \mathcal{N}(0, I)$. See Appendix E for details.

Topological supervision. Following the principle of promoting desired features and suppressing the noisy ones in topology-aware generation Gupta et al. (2025), we complement the CVAE objective with a spatially localized MMA loss.

Each prescribed object is associated with an individual Gaussian distribution g_j and a label $h_j \in \{0, 1\}$. These labels are the respective of the two conditioning channels of the model. For each Gaussian separately, we construct the bifiltration $f_j = (1 - \hat{x},\, 1 - g_j)$ where g_j is normalized. The loss $\mathcal{L}_{\text{individual}}$ is therefore a sum of per-Gaussian contributions, rather than a single evaluation on either channel's combined map. Each contribution encourages one connected component at the corresponding location and, when $h_j = 1$, one additional degree-one feature representing a hole. The full Gaussian mixture is used separately to compute $\mathcal{L}_{\text{outside}}$.

Let $A_{j,i}^{H_q}$ denote the area of the i-th MMA summand support in degree q, cropped to $[0, 1]^2$. To downweight late-born features continuously, define

$$\omega(b) = \max\left\{0,\, \min\left\{1,\, \frac{\tau - b}{\Delta}\right\}\right\}, \qquad \Delta > 0. \tag{3}$$

Here, τ is the birth threshold and Δ is a fixed hyperparameter specifying the width of the transition interval $[\tau - \Delta, \tau]$. We use $\tau = 0.9$ and $\Delta = 0.1$. The adjusted areas are

$$\widetilde{A}_{j,i}^{H_q} = A_{j,i}^{H_q}\, \omega\Big(b_{j,i}^{H_q}\Big), \tag{4}$$

where $b_{j,i}^{H_q}$ is the smallest finite first coordinate among the summand's birth corners. If no such coordinate exists, the weight is set to one.

The weight ω is $1/\Delta$-Lipschitz, and all summands are retained without hard size thresholds or truncation. Each term of $\mathcal{L}_{\text{topo}}$ is computed on the MMA of one bifiltration and satisfies condition (3) of Theorem 12, which therefore applies to each term, and its conclusions pass to the sum at every parameter where all the bifiltrations satisfy its hypotheses.

Individual topology. To compute the individual loss we order the adjusted areas in decreasing order, taking the largest area to be zero for an empty family. Let $h_j = 1$ for a hollow object and $h_j = 0$ for a filled object. We define

$$\mathcal{L}_{\text{individual}} = \sum_j \Big[-w_0 \widetilde{A}_{j,1}^{H_0} + \alpha \sum_{i \geq 2} \widetilde{A}_{j,i}^{H_0} + h_j\big(-w_1 \widetilde{A}_{j,1}^{H_1} + \alpha \sum_{i \geq 2} \widetilde{A}_{j,i}^{H_1}\big) + (1 - h_j)\, \alpha \sum_i \widetilde{A}_{j,i}^{H_1}\Big]. \tag{5}$$

This encourages one dominant connected component per object and, for hollow objects, one dominant hole. All remaining adjusted areas are penalized. The coefficients $w_0, w_1 > 0$ control the rewards for desired features, while $\alpha \geq 0$ controls the penalty on additional features.

Outside topology. Let g_{mix} denote the normalized complete mixture of Gaussians. We construct $f_{\text{out}} = (1 - \hat{x},\, g_{\text{mix}})$, so that low-density regions enter early along the spatial coordinate, and define

$$\mathcal{L}_{\text{outside}} = \lambda_{\text{out}} \sum_i A_{\text{out},i}^{H_0}. \tag{6}$$

This penalizes all degree-zero summand areas without applying the late-birth weight. It is computed once per image, independently of the number of objects. The complete topological loss is

$$\mathcal{L}_{\text{topo}} = \mathcal{L}_{\text{individual}} + \mathcal{L}_{\text{outside}}. \tag{7}$$

Training objective. We apply the topological loss both to posterior reconstructions and to prior samples, which is the regime used at inference:

$$\begin{aligned}\mathcal{L}_{\text{total}}(x, c) = \mathbb{E}_{z \sim q_\phi(z|x,c)}\big[\mathcal{L}_{\text{BCE}}(x, \hat{x}_z) + \lambda_{\text{topo}}\, \mathcal{L}_{\text{topo}}(\hat{x}_z; c)\big] + \lambda_{\text{gen}}\, \mathbb{E}_{z \sim \mathcal{N}(0,I)}\big[\mathcal{L}_{\text{topo}}(\hat{x}_z; c)\big] \\ + \beta_t\, D_{\text{KL}}\big(q_\phi(z \mid x, c)\,\|\,\mathcal{N}(0, I)\big), \qquad \hat{x}_z = D_\theta(z, c).\end{aligned} \tag{8}$$

Here, $\mathcal{L}_{\text{BCE}}$ is binary cross-entropy summed over pixels, and the dependence of $\mathcal{L}_{\text{topo}}$ on c denotes the associated spatial distributions and object labels. Each expectation is estimated with one sample per image. Losses, minibatch and coefficient schedules are detailed in Appendix E.

Method ↓	$N = 1$	$N = 2$	$N = 3$	$N = 4$	$N = 5$	$N = 6$	$N = 7$	$N = 8$	$N = 9$	$N = 10$
MMA	0.04 ± 0.02	0.04 ± 0.01	0.04 ± 0.02	0.04 ± 0.02	0.05 ± 0.05	0.05 ± 0.03	0.06 ± 0.05	0.08 ± 0.08	0.10 ± 0.11	0.11 ± 0.10
1PH	0.09 ± 0.20	0.11 ± 0.20	0.12 ± 0.17	0.09 ± 0.10	0.12 ± 0.15	0.10 ± 0.10	0.09 ± 0.08	0.09 ± 0.07	0.09 ± 0.07	0.10 ± 0.08
Ablation	0.53 ± 0.86	0.46 ± 0.51	0.41 ± 0.46	0.50 ± 0.39	0.45 ± 0.29	0.53 ± 0.32	0.56 ± 0.34	0.57 ± 0.27	0.59 ± 0.29	0.62 ± 0.29

Table 1: Mean benchmark scores $\pm$ standard deviation over 100 test images per object count N (lower is better). Counts $N \leq 5$ are in-distribution; $N \geq 6$ are out-of-distribution. Ablation denotes training without a topological loss.

4.2 Evaluation and Results

Dataset. We use synthetic binary masks containing randomly filled or hollow triangles, squares, and circles, each associated with a Gaussian spatial distribution and a topology label. The training set contains $2{,}500$ images, with 500 per object count from one to five; the test set contains $1{,}000$ images, with 100 per count from one to ten. Counts one to five are in-distribution and six to ten are out-of-distribution, with unchanged shape families and topological classes. We report results by subset and object count to assess generalization to more complex scenes in Table 1.

Metric. A generated mask is correct when every prescribed location hosts one object of the prescribed topology and no object appears elsewhere; global Betti numbers ignore locations, and pixel-wise accuracy penalizes shape variations that the conditioning leaves free. We therefore detect holes and filled components in the thresholded mask, match them to the target centers, and charge each missing or spurious object 1, each object of the wrong homological degree 0.5, and each correctly detected object less than 0.5 for its displacement and its MMA quality deficit. The image score S is the total cost divided by the number of targets; lower is better. See Appendix F for more details.

Results. Table 1 compares CVAEs trained with MMA supervision, single-parameter persistent homology (1PH), and no topological loss. MMA-based models achieves the lowest mean benchmark score at every in-distribution object count ($N = 1, \ldots, 5$), with means between 0.04 and 0.05, compared with 0.09–0.12 for 1PH and 0.41–0.53 for the ablation. Its scores also exhibit lower variability across test images than either baseline within this range. For unseen object counts, MMA-based models retain the lowest mean score at $N = 6, 7, 8$, but its error increases as scenes become more crowded. At $N = 9, 10$, 1PH obtains slightly lower mean scores, indicating that MMA's advantage does not extend uniformly to all tested object counts. Nevertheless, using MMA substantially improves upon the non-topological ablation at every count. The results therefore demonstrate improved conditional generation within the training range and for moderately larger scenes, while highlighting a limitation when generalizing to the largest unseen object counts.

5 Conclusions

We presented a framework for incorporating multi-parameter PH into gradient-based learning through Multi-parameter Module Approximations, and applied it to constrain image generation with prescribed local topology. We showed that expressing MMA through corner coordinates on combinatorial blocks yields regularity properties under explicit genericity assumptions, conditions under which the resulting objectives are definable and locally Lipschitz, and stochastic subgradient convergence guarantees in the fixed-block setting. Building on this, we introduced a localized topological loss that combines image intensity and spatial reference distributions in a bifiltration, and so ties connected components and holes to prescribed locations while penalizing unwanted features. On our benchmark experiments, MMA outperforms both single-parameter topological supervision and the ablation without topological loss.

As for future work, our guarantees hold at generic filtrations and along the segments described by Assumptions 8 and 9, and characterizing the block boundaries where they fail is the natural next step. Moreover, there are many tricks in single-parameter PH that can be used to speed up computations and ensure stronger theoretical guarantees than convergence (see Carrière et al. (2026)); whether these can be adapted to multi-parameter PH and MMA is an open question.

AI USE STATEMENT

In this work, we used generative AI tools to implement methods, helping write the code for our proposed approach, and to assist in the writing of proofs, by checking our proofs for errors and gaps and suggesting corrections. We have not used generative AI tools to formulate mathematical claims, develop theoretical models or conceptual frameworks, provide critical ingredients for proving mathematical claims, propose or refine hypotheses, design or provide feedback on research methodology or experiments, or interpret results. Generating synthetic data sets, cleaning or reformatting datasets, translation, and qualitative or thematic data analysis are not applicable to this work. Additionally, we used generative AI tools to create and edit auxiliary software code, and to draft and edit parts of the paper to improve readability. We have reviewed all AI-assisted work. The research ideas, theoretical results and proof strategies are the authors' own. We adopted AI-suggested proof corrections only after independently verifying them, and we checked every proof in full ourselves. We manually reviewed and tested all AI-assisted code for correctness, and we read and revised all AI-assisted text. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work uses only synthetic binary masks generated by the authors, and raises no concerns regarding human subjects, personal data, or dataset licensing.

REPRODUCIBILITY STATEMENT

An anonymous repository containing the complete downloadable dataset, our model, and the scripts necessary to reproduce our results is provided at `https://anonymous.4open.science/r/ICLR_Submission-D768`

REFERENCES

Rubén Ballester, Carles Casacuberta, and Sergio Escalera. *Topological Data Analysis for Neural Networks*. SpringerBriefs in Computer Science. Springer Nature Switzerland, Cham, 2026. ISBN 978-3-032-08282-4 978-3-032-08283-1. doi: 10.1007/978-3-032-08283-1.

Serguei Barannikov, Daria Voronkova, Alexander Mironenko, Ilya Trofimov, Alexander Korotin, Grigorii Sotnikov, and Evgeny Burnaev. Loss Barcode: A Topological Measure of Escapability in Loss Landscapes. *Doklady Mathematics*, 112(3):600–623, December 2025. ISSN 1531-8362. doi: 10.1134/S106456242570053X.

Matteo Biagetti, Juan Calles, Lina Castiblanco, Alex Cole, and Jorge Noreña. Fisher Forecasts for Primordial Non-Gaussianity from Persistent Homology. *Journal of Cosmology and Astroparticle Physics (JCAP)*, 2022(10):002, October 2022. ISSN 1475-7516. doi: 10.1088/1475-7516/2022/10/002.

Lital Binyamin, Yoad Tewel, Hilit Segev, Eran Hirsch, Royi Rassin, and Gal Chechik. Make it count: Text-to-image generation with an accurate number of objects. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13242–13251, June 2025.

Magnus Bakke Botnan and Michael Lesnick. An Introduction to Multiparameter Persistence. In Aslak Bakke Buan, Henning Krause, and Øyvind Solberg (eds.), *Representations of Algebras and Related Structures*, pp. 77–150. European Mathematical Society, November 2023. ISBN 978-3-98547-054-9 978-3-98547-554-4. doi: 10.4171/ecr/19/4.

Gunnar Carlsson and Afra Zomorodian. The Theory of Multidimensional Persistence. *Discrete & Computational Geometry (DCG)*, 42(1):71–93, July 2009. ISSN 1432-0444. doi: 10.1007/s00454-009-9176-0.

Mathieu Carrière, Steve Oudot, and Maks Ovsjanikov. Stable Topological Signatures for Points on 3D Shapes. *Computer Graphics Forum*, 34(5):1–12, 2015. doi: 10.1111/cgf.12692.

Mathieu Carrière, Frédéric Chazal, Yuichi Ike, Théo Lacombe, Martin Royer, and Yuhei Umeda. PersLay: A Neural Network Layer for Persistence Diagrams and New Graph Topological Signatures. In Silvia Chiappa and Roberto Calandra (eds.), *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics (AISTATS 2020)*, volume 108, pp. 2786–2796. PMLR, June 2020.

Mathieu Carrière, Frédéric Chazal, Marc Glisse, Yuichi Ike, Hariprasad Kannan, and Yuhei Umeda. Optimizing Persistent Homology Based Functions. In Marina Meila and Tong Zhang (eds.), *Proceedings of the 38th International Conference on Machine Learning (ICML 2021)*, volume 139, pp. 1294–1303. PMLR, July 2021.

Mathieu Carrière, Marc Theveneau, and Théo Lacombe. Diffeomorphic Interpolation for Efficient Persistence-Based Topological Optimization. In Amir Globerson, Lester Mackey, Danielle Belgrave, Angela Fan, Ulrich Paquet, Jakub Tomczak, and Cheng Zhang (eds.), *Advances in Neural Information Processing Systems 37 (NeurIPS 2024)*, volume 37, pp. 27274–27294. Curran Associates, Inc., 2024.

Mathieu Carrière, Seunghyun Kim, and Woojin Kim. Sparsification of the Generalized Persistence Diagrams for Scalability Through Gradient Descent. In Oswin Aichholzer and Haitao Wang (eds.), *41st International Symposium on Computational Geometry (SoCG 2025)*, volume 332 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pp. 29:1–29:17, Dagstuhl, Germany, 2025. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. ISBN 978-3-95977-370-6. doi: 10.4230/LIPIcs.SoCG.2025.29.

Mathieu Carrière, Yuichi Ike, Théo Lacombe, and Naoki Nishikawa. Persistence-Based Topological Optimization: A Survey, March 2026.

Frédéric Chazal and Bertrand Michel. An Introduction to Topological Data Analysis: Fundamental and Practical Aspects for Data Scientists. *Frontiers in Artificial Intelligence*, 4, September 2021. ISSN 2624-8212. doi: 10.3389/frai.2021.667963.

Frédéric Chazal, Clément Levrard, and Martin Royer. Topological Analysis for Detecting Anomalies (TADA) in Dependent Sequences: Application to Time Series. *Journal of Machine Learning Research (JMLR)*, 25(365):1–49, 2024. ISSN 1533-7928.

Chao Chen, Xiuyan Ni, Qinxun Bai, and Yusu Wang. A Topological Regularizer for Classifiers via Persistent Homology. In Kamalika Chaudhuri and Masashi Sugiyama (eds.), *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS 2019)*, volume 89, pp. 2573–2582. PMLR, April 2019.

James R. Clough, Nicholas Byrne, Ilkay Oksuz, Veronika A. Zimmer, Julia A. Schnabel, and Andrew P. King. A topological loss function for deep-learning based image segmentation using persistent homology. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44 (12):8766–8778, December 2022. ISSN 1939-3539. doi: 10.1109/tpami.2020.3013679. URL http://dx.doi.org/10.1109/TPAMI.2020.3013679.

David Cohen-Steiner, Herbert Edelsbrunner, and Dmitriy Morozov. Vines and vineyards by updating persistence in linear time. In *Proceedings of the Twenty-Second Annual Symposium on Computational Geometry*, SCG '06, pp. 119–126, New York, NY, USA, 2006. Association for Computing Machinery. ISBN 1595933409. doi: 10.1145/1137856.1137877. URL https://doi.org/10.1145/1137856.1137877.

William Crawley-Boevey. Decomposition of pointwise finite-dimensional persistence modules. *Journal of Algebra and Its Applications*, 14(05):1550066, 2015. doi: 10.1142/S0219498815500668. URL https://doi.org/10.1142/S0219498815500668.

Damek Davis, Dmitriy Drusvyatskiy, Sham Kakade, and Jason D Lee. Stochastic subgradient method converges on tame functions. *Foundations of computational mathematics*, 20(1):119–154, 2020.

Tamal K. Dey and Yusu Wang. *Computational Topology for Data Analysis*. Cambridge University Press, 2022. ISBN 978-1-009-09816-8. doi: 10.1017/9781009099950.

Tamal K. Dey and Cheng Xin. Computing Bottleneck Distance for 2-D Interval Decomposable Modules. In Bettina Speckmann and Csaba D. Tóth (eds.), *34th International Symposium on Computational Geometry (SoCG 2018)*, volume 99 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pp. 32:1–32:15, Dagstuhl, Germany, 2018. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. ISBN 978-3-95977-066-8. doi: 10.4230/LIPIcs.SoCG.2018.32. URL https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.SoCG.2018.32.

S. M. Ali Eslami, Nicolas Heess, Theophane Weber, Yuval Tassa, David Szepesvari, koray kavukcuoglu, and Geoffrey E Hinton. Attend, infer, repeat: Fast scene understanding with generative models. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016. URL https://proceedings.neurips.cc/paper_files/paper/2016/file/52947e0ade57a09e4a1386d08f17b656-Paper.pdf.

Saumya Gupta, Dimitris Samaras, and Chao Chen. TopoDiffusionNet: A Topology-Aware Diffusion Model. In *The 13th International Conference on Learning Representations (ICLR 2025)*. OpenReview.net, 2025.

Olympio Hacquard, Krishnakumar Balasubramanian, Gilles Blanchard, Clément Levrard, and Wolfgang Polonik. Topologically Penalized Regression on Manifolds. *Journal of Machine Learning Research (JMLR)*, 23(161):1–39, 2022. ISSN 1533-7928.

Christoph Hofer, Florian Graf, Bastian Rieck, Marc Niethammer, and Roland Kwitt. Graph Filtration Learning. In Hal Daumé III and Aarti Singh (eds.), *Proceedings of the 37th International Conference on Machine Learning (ICML 2020)*, volume 119, pp. 4314–4323. PMLR, November 2020.

Max Horn, Edward de Brouwer, Michael Moor, Yves Moreau, Bastian Rieck, and Karsten M. Borgwardt. Topological Graph Neural Networks. In *The 10th International Conference on Learning Representations (ICLR 2022)*. OpenReview.net, 2022.

Elena K. Kandror, Mathieu Carrière, Alexis Peterson, Will Liao, Andreas Tjärnberg, Jun Hou Fung, Krishnaa T. Mahbubani, Jackson Loper, William Pangburn, Yuchen Xu, Kourosh Saeb-Parsy, Tom Maniatis, and Abbas H. Rizvi. Enhancer Dynamics and Cellular Architecture in the Human Spinal Cord. *Neuron*, 114(8):1399–1418.e18, April 2026. ISSN 0896-6273. doi: 10.1016/j.neuron.2025.12.035.

Michael Kerber and Florian Russold. Representing Two-Parameter Persistence Modules via Graphcodes, March 2025.

Caner Korkmaz, Brighton Nuwagira, Barış Coşkunuzer, and Tolga Birdal. CuMPerLay: Learning Cubical Multiparameter Persistence Vectorizations, October 2025.

Manik Kuchroo, Marcello DiStasio, Eric Song, Eda Calapkulu, Le Zhang, Maryam Ige, Amar H. Sheth, Abdelilah Majdoubi, Madhvi Menon, Alexander Tong, Abhinav Godavarthi, Yu Xing, Scott Gigante, Holly Steach, Jessie Huang, Guillaume Huguet, Janhavi Narain, Kisung You, George Mourgkos, Rahul M. Dhodapkar, Matthew J. Hirn, Bastian Rieck, Guy Wolf, Smita Krishnaswamy, and Brian P. Hafler. Single-Cell Analysis Reveals Inflammatory Interactions Driving Macular Degeneration. *Nature Communications*, 14(1):2589, May 2023. ISSN 2041-1723. doi: 10.1038/s41467-023-37025-7.

Théo Lacombe, Yuichi Ike, Mathieu Carrière, Frédéric Chazal, Marc Glisse, and Yuhei Umeda. Topological Uncertainty: Monitoring Trained Neural Networks through Persistence of Activation Graphs. In Zhi-Hua Zhou (ed.), *30th International Joint Conference on Artificial Intelligence (IJCAI 2021)*, pp. 2666–2672. International Joint Conferences on Artificial Intelligence Organization, August 2021. doi: 10.24963/ijcai.2021/367.

Claudia Landi. *The Rank Invariant Stability via Interleavings*, pp. 1–10. Springer International Publishing, Cham, 2018. ISBN 978-3-319-89593-2. doi: 10.1007/978-3-319-89593-2_1. URL https://doi.org/10.1007/978-3-319-89593-2_1.

Michael Lesnick. The theory of the interleaving distance on multidimensional persistence modules. *Foundations of Computational Mathematics*, 15(3):613–650, 2015.

Michael Lesnick and Matthew Wright. Interactive visualization of 2-d persistence modules, 2015. URL https://arxiv.org/abs/1512.00180.

Jacob Leygonie, Mathieu Carrière, Théo Lacombe, and Steve Oudot. A Gradient Sampling Algorithm for Stratified Maps with Applications to Topological Data Analysis. *Mathematical Programming*, 202(1):199–239, November 2023. ISSN 1436-4646. doi: 10.1007/s10107-023-01931-x.

Yuheng Li, Haotian Liu, Qingyang Wu, Fangzhou Mu, Jianwei Yang, Jianfeng Gao, Chunyuan Li, and Yong Jae Lee. Gligen: Open-set grounded text-to-image generation. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 22511–22521, 2023. doi: 10.1109/CVPR52729.2023.02156.

Rosanne Liu, Joel Lehman, Piero Molino, Felipe Petroski Such, Eric Frank, Alex Sergeev, and Jason Yosinski. An intriguing failing of convolutional neural networks and the coordconv solution. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/file/60106888f8977b71e1f15db7bc9a88d1-Paper.pdf.

David Loiseaux, Mathieu Carrière, and Andrew J. Blumberg. A Framework for Fast and Stable Representations of Multiparameter Persistent Homology Decompositions. In Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine (eds.), *Advances in Neural Information Processing Systems 36 (NeurIPS 2023)*, volume 36, pp. 35774–35798. Curran Associates, Inc., 2023a.

David Loiseaux, Luis Scoccola, Mathieu Carrière, Magnus Bakke Botnan, and Steve Oudot. Stable Vectorization of Multiparameter Persistent Homology using Signed Barcodes as Measures. In Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine (eds.), *Advances in Neural Information Processing Systems 36 (NeurIPS 2023)*, volume 36, pp. 68316–68342. Curran Associates, Inc., 2023b.

David Loiseaux, Mathieu Carrière, and Andrew J Blumberg. Multi-parameter module approximation: an efficient and interpretable invariant for multi-parameter persistence modules with guarantees. *Journal of Applied and Computational Topology*, 9(4):1–60, 2025.

Michael Moor, Max Horn, Bastian Rieck, and Karsten Borgwardt. Topological Autoencoders. In Hal Daumé III and Aarti Singh (eds.), *Proceedings of the 37th International Conference on Machine Learning (ICML 2020)*, volume 119, pp. 7045–7054. PMLR, November 2020.

Soham Mukherjee, Shreyas N. Samaga, Cheng Xin, Steve Oudot, and Tamal K. Dey. D-GRIL: End-To-End Topological Learning with 2-Parameter Persistence. In Hee-Kap Ahn, Michael Hoffmann, and Amir Nayyeri (eds.), *42nd International Symposium on Computational Geometry (SoCG 2026)*, volume 367 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pp. 79:1–79:17, Dagstuhl, Germany, 2026. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. ISBN 978-3-95977-418-5. doi: 10.4230/LIPIcs.SoCG.2026.79.

Maxime Multari, Mathieu Carrière, Xavier Amorós-Gabarrón, Krishna Kumar, Pratila Debnath, Alexina Damy, Ludo Andrianirina Mamisoa, Sebastian Lobentanzer, Julio Saez-Rodriguez, Stéphanie Jaubert, Aurélien Dugourd, Justyna J. Olas, and Silvia Bottini. From Knowledge Graph to Topological Data Analysis: A Novel Framework to Analyze Gene Regulatory Networks for Tomato – Multi-Pathogen Interactions. *New Phytologist*, 2026. ISSN 1469-8137. doi: 10.1111/nph.71382.

Arnur Nigmetov and Dmitriy Morozov. Topological Optimization with Big Steps. *Discrete & Computational Geometry (DCG)*, 72(1):310–344, July 2024. ISSN 1432-0444. doi: 10.1007/s00454-023-00613-x.

Doruk Oner, Adélie Garin, Mateusz Koziński, Kathryn Hess, and Pascal Fua. Persistent homology with improved locality information for more effective delineation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(8):10588–10595, 2023. doi: 10.1109/TPAMI.2023.3246921.

Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Semantic image synthesis with spatially-adaptive normalization. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2332–2341, 2019. doi: 10.1109/CVPR.2019.00244.

Jose A. Perea and John L. Harer. Sliding Windows and Persistence: An Application of Topological Methods to Signal Analysis. *Foundations of Computational Mathematics (FoCM)*, 15(3):799–838, June 2015. ISSN 1615-3383. doi: 10.1007/s10208-014-9206-z.

Adrien Poulenard, Primoz Skraba, and Maks Ovsjanikov. Topological Function Optimization for Continuous Shape Matching. *Computer Graphics Forum*, 37(5):13–25, 2018. ISSN 1467-8659. doi: 10.1111/cgf.13487.

Bastian Rieck, Matteo Togninalli, Christian Bock, Michael Moor, Max Horn, Thomas Gumbsch, and Karsten M. Borgwardt. Neural Persistence: A Complexity Measure for Deep Neural Networks using Algebraic Topology. In *The 7th International Conference on Learning Representations (ICLR 2019)*. OpenReview.net, 2019.

Luis Scoccola, Siddharth Setlur, David Loiseaux, Mathieu Carrière, and Steve Oudot. Differentiability and Optimization of Multiparameter Persistent Homology. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning (ICML 2024)*, pp. 43986–44011. PMLR, July 2024.

Primoz Skraba, Maks Ovsjanikov, Frédéric Chazal, and Leonidas J. Guibas. Persistence-Based Segmentation of Deformable Shapes. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW 2010)*, pp. 45–52. IEEE, June 2010. doi: 10.1109/CVPRW.2010.5543285.

Kihyuk Sohn, Honglak Lee, and Xinchen Yan. Learning structured output representation using deep conditional generative models. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015. URL https://proceedings.neurips.cc/paper_files/paper/2015/file/8d55a249e6baa5c06772297520da2051-Paper.pdf.

Lou Van den Dries. *Tame topology and o-minimal structures*, volume 248. Cambridge university press, 1998.

Lou van den Dries and Chris Miller. Geometric categories and o-minimal structures. *Duke Mathematical Journal*, 84:497–540, 1996. URL https://api.semanticscholar.org/CorpusID:274110.

Robin Vandaele, Bo Kang, Jefrey Lijffijt, Tijl de Bie, and Yvan Saeys. Topologically Regularized Data Embeddings. In *The 10th International Conference on Learning Representations (ICLR 2022)*. OpenReview.net, 2022.

Oliver Vipond. Multiparameter Persistence Landscapes. *Journal of Machine Learning Research (JMLR)*, 21(61):1–38, 2020. ISSN 1533-7928.

Fan Wang, Huidong Liu, Dimitris Samaras, and Chao Chen. TopoGAN: A Topology-Aware Generative Adversarial Network. In *16th European Conference on Computer Vision (ECCV 2020)*, pp. 118–136, Berlin, Heidelberg, August 2020. Springer-Verlag. ISBN 978-3-030-58579-2. doi: 10.1007/978-3-030-58580-8_8.

Nick Watters, Loic Matthey, Chris P. Burgess, and Alexander Lerchner. Spatial broadcast decoder: A simple architecture for disentangled representations in VAEs, 2019. URL https://openreview.net/forum?id=S1x7WjnzdV.

Jingjie Yang, Heidi Fang, Jagdeep Dhesi, Iris H. R. Yoon, Joshua A. Bull, Helen M. Byrne, Heather A. Harrington, and Gillian Grindstaff. Topological Classification of Tumour-Immune Interactions and Dynamics. *Journal of Mathematical Biology*, 91(3):25, August 2025. ISSN 1432-1416. doi: 10.1007/s00285-025-02253-6.

Jacky H. T. Yip, Matteo Biagetti, Alex Cole, Karthik Viswanathan, and Gary Shiu. Cosmology with Persistent Homology: A Fisher Forecast. *Journal of Cosmology and Astroparticle Physics (JCAP)*, 2024(9):034, September 2024. ISSN 1475-7516. doi: 10.1088/1475-7516/2024/09/034.

Jacky H. T. Yip, Adam Rouhiainen, and Gary Shiu. Learning from Topology: Cosmological Parameter Estimation from the Large-Scale Structure. *Machine Learning: Science and Technology*, 6 (2):025063, June 2025. ISSN 2632-2153. doi: 10.1088/2632-2153/ade114.

Sebastian Zeng, Florian Graf, Christoph Hofer, and Roland Kwitt. Topological Attention for Time Series Forecasting. In Marc'Aurelio Ranzato, Alina Beygelzimer, Yann Dauphin, Percy S. Liang, and Jenn Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems 34 (NeurIPS 2021)*, volume 34, pp. 24871–24882. Curran Associates, Inc., 2021.

Sebastian Zeng, Florian Graf, Martin Uray, Stefan Huber, and Roland Kwitt. Neural Persistence Dynamics. In Amir Globerson, Lester Mackey, Danielle Belgrave, Angela Fan, Ulrich Paquet, Jakub Tomczak, and Cheng Zhang (eds.), *Advances in Neural Information Processing Systems 37 (NeurIPS 2024)*, volume 37, pp. 33286–33309. Curran Associates, Inc., 2024.

Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 3836–3847, October 2023.

A Additional background on persistent homology

This appendix expands the material compressed in Section 2. It follows Lesnick (2015); see Carlsson & Zomorodian (2009) for the foundational theory and Botnan & Lesnick (2023) for a survey.

A.1 Persistence modules as graded modules

Section 2 describes a persistence module by its vector spaces and transition maps. There is an equivalent algebraic description, and it is the one carrying the shift operation used to define the interleaving distance. Let P_n be the polynomial ring $\Bbbk[x_1, \dots, x_n]$, where exponents of the indeterminates in P_n take arbitrary values in $[0, \infty)$. An *n-graded module* (or simply a n-module) is a P_n-module $\mathbb{M}$ with a direct sum decomposition as a $\Bbbk$-vector space $\mathbb{M} \simeq \bigoplus_{a \in \mathbb{R}^n} M_a$, such that $x^b(M_a) \subset M_{a+b}$ for all $a \in \mathbb{R}^n$, $b \in [0, \infty)^n$. For $a \leq_n b$, the action of x^{b-a} on $\mathbb{M}$ defines a linear map $M_a \to M_b$, which we denote by $\varphi_{\mathbb{M}}(a, b)$ and call a *transition map*. We define a category n-$\mathbf{Mod}$ of n-modules whose morphisms $f \colon \mathbb{M} \to \mathbb{M}'$ are grading-preserving P_n-module homomorphisms, i.e., $f(M_a) \subseteq M'_a$ for all $a \in \mathbb{R}^n$. Writing $f_a \colon M_a \to M'_a$ for the restriction, functoriality gives $f_b \circ \varphi_{\mathbb{M}}(a, b) = \varphi_{\mathbb{M}'}(a, b) \circ f_a$ for all $a \leq_n b$. That is, the following diagram commutes for all $a \leq_n b \in \mathbb{R}^n$:

$$\begin{array}{ccc} M_a & \xrightarrow{\varphi_{\mathbb{M}}(a,b)} & M_b \\ {\scriptstyle f_a}\downarrow & & \downarrow{\scriptstyle f_b} \\ M'_a & \xrightarrow{\varphi_{\mathbb{M}'}(a,b)} & M'_b \end{array} .$$

A.2 The interleaving distance

The n-$\mathbf{Mod}$ interpretation allows us to define a metric between persistence modules. For $v \in \mathbb{R}^n$, the *v-shift functor* $(\cdot)(v) \colon n\text{-}\mathbf{Mod} \to n\text{-}\mathbf{Mod}$ is defined by $M(v)_a = M_{a+v}$ and $\varphi_{\mathbb{M}(v)}(a, b) = \varphi_{\mathbb{M}}(a + v, b + v)$ for $a \leq_n b \in \mathbb{R}^n$ and $\mathbb{M}$ a n-module. For a morphism f in n-$\mathbf{Mod}$, $f(v)$ is defined by $f(v)_a = f_{a+v}$. For $t \in \mathbb{R}$, we write $\mathbf{t} = (t, \dots, t) \in \mathbb{R}^n$. With a slight abuse of notation, if $\varepsilon \in \mathbb{R}$, we refer to the $(\cdot)(\varepsilon \mathbf{1})$ shift functor simply as $(\cdot)(\varepsilon)$. Given a n-module $\mathbb{M}$ and $\varepsilon \in [0, \infty)$, the *$\varepsilon$-transition morphism* $\varphi^{\varepsilon}_{\mathbb{M}} \colon \mathbb{M} \to \mathbb{M}(\varepsilon)$ has components $(\varphi^{\varepsilon}_{\mathbb{M}})_a = \varphi_{\mathbb{M}}(a,\, a + \varepsilon \mathbf{1})$, for all $a \in \mathbb{R}^n$.

Definition 14. Two persistence modules $\mathbb{M}$ and $\mathbb{M}'$ are *ε-interleaved* if there exist morphisms $f\colon \mathbb{M} \to \mathbb{M}'(\varepsilon)$ and $g\colon \mathbb{M}' \to \mathbb{M}(\varepsilon)$ satisfying $g(\varepsilon) \circ f = \varphi_{\mathbb{M}}^{2\varepsilon}$ and $f(\varepsilon) \circ g = \varphi_{\mathbb{M}'}^{2\varepsilon}$. The **interleaving distance** is

$$d_I(\mathbb{M}, \mathbb{M}') \coloneqq \inf\big\{\varepsilon \geq 0 \mid \mathbb{M} \text{ and } \mathbb{M}' \text{ are } \varepsilon\text{-interleaved}\big\}.$$

Theorem 15 (Theorem 5.3 (Lesnick, 2015))**.** *The interleaving distance d_I is stable. That is, for any $f, g \in \mathrm{Fil}_n(\mathcal{K})$, one has $d_I\big(\mathbb{M}(f), \mathbb{M}(g)\big) \leq \|f - g\|_\infty$.*

A.3 Lines and the fibered barcode

Section 2 fixes L to be a finite set of equally spaced diagonal lines. We now make this precise. A *diagonal line* is a line of the form $l_b \coloneqq b + t\mathbf{1}$, for $t \in \mathbb{R}$ and $b \in \mathbb{R}^{n-1} \times \{0\}$. Two diagonal lines l, l' are *δ-consecutive* if $l' = l \pm \delta v$, for some $v \in \{0,1\}^n \setminus \{\mathbf{0}, \mathbf{1}\}$. A *finite set of equally spaced diagonal lines* with spacing δ is a finite set of δ-consecutive lines L such that $L = \{l_{\delta u} \mid u \in \mathcal{U}\}$ for some finite $\mathcal{U} \subset \mathbb{Z}^{n-1} \times \{0\}$.

A.4 Compatible matchings

The stability estimate quoted in Section 2 holds for matchings that pair only bars whose endpoints are not strictly comparable.

Definition 16. Given a persistence module $\mathbb{M}$ and $l_1, l_2 \in L$ at distance δ, assume that $\mathrm{supp}(\mathbb{M}) \cap l_1$ and $\mathrm{supp}(\mathbb{M}) \cap l_2$ are not empty, and let $[b_1, d_1)$ and $[b_2, d_2)$ be bars in $\mathcal{B}(M|_{l_1})$ and $\mathcal{B}(M|_{l_2})$, respectively. We say that two bars are **compatible** if their birthpoints (resp. deathpoints) are not strictly comparable, i.e., $b_1 \not<_n b_2$ and $b_2 \not<_n b_1$ (resp. $d_1 \not<_n d_2$ and $d_2 \not<_n d_1$). Moreover, we say that a bar $[b, d) \in l_1$ is *compatible* with the empty set in l_2 if $\|b - d\|_\infty \leq 2\delta$. A **compatible matching function** is a matching function that only pairs compatible bars.

For $n = 2$ the vineyard matching is compatible under Assumption 9, by Remark 32 and Loiseaux et al. (2025, Prop. 17), so the estimate applies to the descriptor used throughout this paper without an explicit check.

A.5 On the notion of interval

The definition of interval adopted in Section 2 is weaker than the one of Loiseaux et al. (2025), which also requires connectedness. A support may therefore be a disjoint union of rectangles, and an arbitrarily small shift of a single corner is enough to disconnect one, as in Figure 3.

We adopt the weaker notion because the losses of Section 4 read the summands produced by the vineyard matching. Under the weaker notion each matched family of bars gives one summand, whose corners depend continuously on the filtration inside a block, and the number of summands is constant on each block. Under the stronger notion the shift of Figure 3 splits one summand into two within a single block, so a loss that rewards the largest summand and penalises the others, as in Equation (5), jumps although the interleaving distance stays small.

B Definability and stochastic subgradient descent

This appendix collects the material compressed in the corresponding paragraph of Section 2, following Davis et al. (2020) and Van den Dries (1998).

Let $F\colon \mathbb{R}^d \to \mathbb{R}$ be locally Lipschitz. By Rademacher's theorem F is differentiable almost everywhere, and its *Clarke subdifferential* at $x \in \mathbb{R}^d$ is

$$\partial F(x) \coloneqq \mathrm{conv}\Big\{ \lim_{k\to\infty} \nabla F(x_k) \;\Big|\; x_k \to x,\ F \text{ differentiable at } x_k \Big\},$$

and x is *critical* for F if $0 \in \partial F(x)$. The *stochastic subgradient method* iterates

$$x_{n+1} = x_n - \alpha_n(y_n + \xi_n), \qquad y_n \in \partial F(x_n),$$

where $\{\alpha_n\}_{n\geq 1}$ is a step-size sequence (the learning rate) and $\{\xi_n\}_{n\geq 1}$ is a sequence of random variables on some probability space.

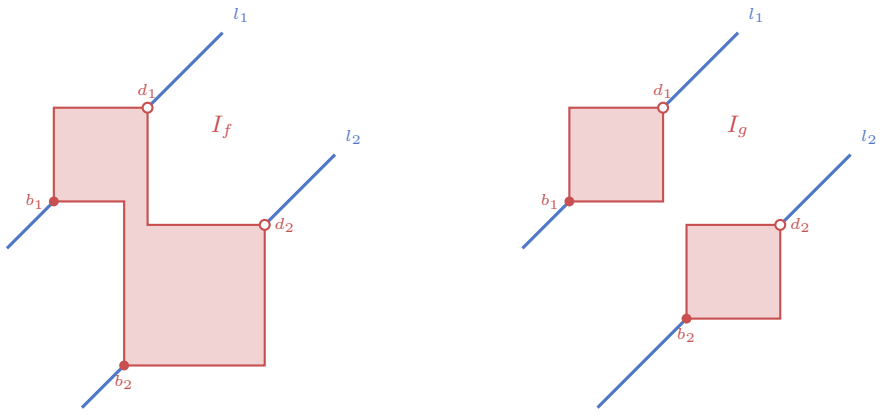


Figure 3: Two supports built from one bar on l_1, with corners b_1 and d_1, and one bar on l_2, with corners b_2 and d_2; the two differ only in the position of b_2 along l_2. On the left $b_2 \leq_2 d_1$ and the support is connected, so I_f is one interval; on the right b_2 has slid past d_1 and the support splits, so I_g is a direct sum of two. The shift is arbitrarily small, and so is the interleaving distance between the two modules, yet under the stronger notion of interval I_f would be one summand and I_g two.

Assumption 17. 1. The learning rate $\{\alpha_n\}_{n\geq 1}$ is nonnegative, square summable, but not summable:

$$\alpha_n \geq 0 \quad \forall n \in \mathbb{N}, \qquad \sum_{n=1}^{\infty} \alpha_n = \infty, \qquad \sum_{n=1}^{\infty} \alpha_n^2 < \infty.$$

2. Almost surely, the iterates are bounded: $\sup_{n\geq 1} \|x_n\| < \infty$.

3. $\{\xi_n\}_{n\geq 1}$ is a martingale difference sequence with respect to the increasing σ-fields $\mathcal{F}_n = \sigma(x_j, y_j, \xi_j : j \leq n)$: there exists $p\colon \mathbb{R}^d \to [0,\infty)$, bounded on bounded sets, such that almost surely, for all $n \in \mathbb{N}$,

$$\mathbb{E}[\xi_n \mid \mathcal{F}_n] = 0 \qquad \text{and} \qquad \mathbb{E}[\|\xi_n\|^2 \mid \mathcal{F}_n] \leq p(x_n).$$

Proposition 18 (Corollary 5.9, Davis et al., 2020)**.** *Let $F\colon \mathbb{R}^d \to \mathbb{R}$ be a locally Lipschitz function that is C^d-stratifiable. Consider the iterates $\{x_n\}_{n\geq 1}$ produced by the stochastic subgradient method and suppose that Assumption 17 holds. Then almost surely, every limit point of $\{x_n\}_{n\geq 1}$ is critical for F, and the sequence of values $\{F(x_n)\}_{n\geq 1}$ converges.*

The class of semiaffine functions is of particular interest because these functions are C^∞-stratifiable (Davis et al., 2020), thus satisfying the key hypothesis of Proposition 18 and ensuring convergence of the stochastic subgradient method to critical points.

Definition 19. A set $S \subseteq \mathbb{R}^n$ is **semiaffine** (or semilinear) if it is a finite union of sets of the form

$$\{x \in \mathbb{R}^n \mid P_1(x) = \cdots = P_p(x) = 0,\ Q_1(x) > 0, \ldots, Q_q(x) > 0\},$$

where P_i, Q_j are affine functions on $\mathbb{R}^n$ for every $i \in [p]$, $j \in [q]$.

Definition 20. A function $F\colon S \to \mathbb{R}^m$ is **semiaffine** if $S \subseteq \mathbb{R}^n$ is semiaffine and the set $\{(x, F(x)) \mid x \in S\} \subseteq \mathbb{R}^n \times \mathbb{R}^m$ is semiaffine.

The definition of semiaffinity extends to any real vector space of finite dimension. Finally, we recall the notion of o-minimal structure. As in the semiaffine setting, any function definable in an o-minimal structure admits a C^p stratification for every $p \geq 1$ (van den Dries & Miller, 1996).

Definition 21. An ***o*-minimal structure** is a sequence of Boolean algebras $\mathcal{O}_n$ of subsets of $\mathbb{R}^n$ such that for every $n \in \mathbb{N}$:

1. if A belongs to $\mathcal{O}_n$, then $A \times \mathbb{R}$ and $\mathbb{R} \times A$ belong to $\mathcal{O}_{n+1}$;

2. if $\pi\colon \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ denotes the coordinate projection onto $\mathbb{R}^n$, then for any A in $\mathcal{O}_{n+1}$ the set $\pi(A)$ belongs to $\mathcal{O}_n$;

3. $\mathcal{O}_n$ contains all sets of the form $\{x \in \mathbb{R}^n \mid p(x) = 0\}$, where p is a polynomial on $\mathbb{R}^n$;
4. the elements of $\mathcal{O}_1$ are exactly the finite unions of intervals (possibly infinite) and points.

The sets A belonging to $\mathcal{O}_n$, for some $n \in \mathbb{N}$, are called **definable** in the o-minimal structure, and so are the maps whose graph is definable.

Remark 22. The finite intersection of semiaffine sets is semiaffine. Moreover, any piecewise affine function is semiaffine. Lastly, semiaffine sets are definable in any o-minimal structure.

C Proofs

C.1 Proofs of Section 3.1

Proof of Proposition 2. Given $f \in \mathrm{Fil}_n(\mathcal{K})$, $\mathrm{block}_L(f)$ is uniquely determined by the tuple of preorders $(\leq_{f,l})_{l \in L}$. Since $\mathcal{K}$ and L are finite, this tuple ranges over a finite set, proving finiteness. As for the partition, $f \in \mathrm{block}_L(f)$ shows that blocks are non-empty and cover $\mathrm{Fil}_n(\mathcal{K})$. To see they are disjoint, suppose $h \in \mathrm{block}_L(f) \cap \mathrm{block}_L(g)$. Then $\leq_{f,l} \equiv \leq_{h,l} \equiv \leq_{g,l}$ for every $l \in L$, hence $f \in \mathrm{block}_L(g)$ and $g \in \mathrm{block}_L(f)$, so $\mathrm{block}_L(f) = \mathrm{block}_L(g)$. □

In what follows, given two simplices $\sigma, \tau \in \mathcal{K}$, we say that τ is *consecutive* to σ w.r.t. a filtration f and a line $l \in L$ if $f^l(\sigma) \leq f^l(\tau)$ and there is no $\varsigma \in \mathcal{K}$, $\varsigma \neq \sigma, \tau$ such that $f^l(\sigma) \leq f^l(\varsigma) \leq f^l(\tau)$.

Proof of Proposition 3. By induction on the number of simplices k. If $k = 1$, each filtration induces the same preorder on any line (the trivial preorder $\sigma \leq_{f,l} \sigma$). Hence, there is only one block, that corresponds to the entire $(\mathbb{R}^n)^{\mathcal{K}}$. Since $(\mathbb{R}^n)^{\mathcal{K}}$ is semiaffine, this concludes the first step. Let us now suppose that blocks are semiaffine if $k = m$, and let us consider a simplicial complex $\mathcal{K}$ with $k = m + 1$ simplices. Let $f \in \mathrm{Fil}_n(\mathcal{K})$ be fixed. We construct a semiaffine set S such that $h \in \mathrm{block}_L(f)$ implies $h \in S$, and $g \notin \mathrm{block}_L(f)$ implies $g \notin S$. This yields $\mathrm{block}_L(f) = S$, concluding the proof. First, let us note that we can write $\mathcal{K} = \mathcal{K}_m \cup \{\sigma\}$, where $\mathcal{K}_m$ is a simplicial complex with m simplices and σ is a facet. By induction hypothesis, blocks of $\mathrm{Fil}_n(\mathcal{K}_m)$ are semiaffine. It holds that $f|_{\mathcal{K}_m} \in \mathrm{Fil}_n(\mathcal{K}_m)$ and $\mathrm{block}_L(f|_{\mathcal{K}_m})$ is semiaffine. Let $\pi \colon (\mathbb{R}^n)^{\mathcal{K}} \to (\mathbb{R}^n)^{\mathcal{K}_m}$, $\pi(h) = h|_{\mathcal{K}_m}$, be the projection forgetting the σ-component, and consider $\pi^{-1}(\mathrm{block}_L(f|_{\mathcal{K}_m})) = \mathrm{block}_L(f|_{\mathcal{K}_m}) \times \mathbb{R}^n$. Since $\mathrm{block}_L(f|_{\mathcal{K}_m})$ is semiaffine by the induction hypothesis and semiaffine sets are stable under product with $\mathbb{R}^n$, the set $\pi^{-1}(\mathrm{block}_L(f|_{\mathcal{K}_m}))$ is semiaffine. Moreover, $\mathrm{block}_L(f) \subseteq \pi^{-1}(\mathrm{block}_L(f|_{\mathcal{K}_m}))$ and if $g \in \mathrm{Fil}_n(\mathcal{K})$ is such that $g|_{\mathcal{K}_m} \notin \mathrm{block}_L(f|_{\mathcal{K}_m})$, then $g \notin \pi^{-1}(\mathrm{block}_L(f|_{\mathcal{K}_m}))$ by definition of π^{-1}. Due to the finiteness of both $\mathcal{K}$ and L, to conclude the proof it suffices to show that, if $f, g \in \mathrm{Fil}_n(\mathcal{K})$ are such that $g|_{\mathcal{K}_m} \in \mathrm{block}_L(f|_{\mathcal{K}_m})$ but $\leq_{f,l} \not\equiv \leq_{g,l}$ for a certain line l (hence $g \notin \mathrm{block}_L(f)$), we can find semiaffine sets $S^{\cup}, S^{\cap}$ such that $g \notin S \coloneqq \mathrm{Fil}_n(\mathcal{K}) \cap \pi^{-1}(\mathrm{block}_L(f|_{\mathcal{K}_m})) \cap S^{\cup} \cap S^{\cap}$, while $h \in S$ for every $h \in \mathrm{block}_L(f)$. By previous claims, $\leq_{f,l}$ and $\leq_{g,l}$ may differ only w.r.t. the facet σ. In particular, one of three things can happen:

1. $f^l(\sigma) \geq f^l(\tau)$ for any $\tau \in \mathcal{K}_m$, but there exists $\varsigma \in \mathcal{K}_m$ such that $g^l(\sigma) \leq g^l(\varsigma)$, with at least one inequality strict;
2. $f^l(\sigma) \leq f^l(\tau)$ for any $\tau \in \mathcal{K}_m$, but there exists $\varsigma \in \mathcal{K}_m$ such that $g^l(\sigma) \geq g^l(\varsigma)$, with at least one inequality strict;
3. there exist consecutive $\tau_1, \tau_2 \in \mathcal{K}_m$ w.r.t. $f|_{\mathcal{K}_m}, l$ such that $f^l(\tau_1) \leq f^l(\sigma) \leq f^l(\tau_2)$, but either $g^l(\sigma) \leq g^l(\tau_1)$ or $g^l(\sigma) \geq g^l(\tau_2)$, with at least one strict inequality between σ and the violated bound.

In what follows, we assume all inequalities are strict; the case of equalities follows by a similar argument, replacing the strict comparisons below by equalities. We first highlight that, for any two simplices $\varsigma_1, \varsigma_2 \in \mathcal{K}$, the set

$$\left\{h \in (\mathbb{R}^n)^{\mathcal{K}} \mid h^l(\varsigma_1) > h^l(\varsigma_2)\right\} = \bigcup_{i=1}^{n} \bigcap_{i'=1}^{n} \left\{h \in (\mathbb{R}^n)^{\mathcal{K}} \mid h(\varsigma_1)_i - b_i^l > h(\varsigma_2)_{i'} - b_{i'}^l\right\}$$

is semiaffine, being a finite union of finite intersections of open affine half-spaces. For ease of notation, let $\bar{\varsigma}, \bar{\varrho} \in \mathcal{K}_m$ be the simplices that achieve, respectively, the maximum and the minimum for f^l over $\mathcal{K}_m$. If (1), we set

$$S^{\cup} = \left\{h \in (\mathbb{R}^n)^{\mathcal{K}} \mid h^l(\sigma) > h^l(\bar{\varsigma})\right\}, \quad S^{\cap} = \mathrm{Fil}_n(\mathcal{K}).$$

If (2), we set

$$S^{\cup} = \mathrm{Fil}_n(\mathcal{K}), \quad S^{\cap} = \left\{h \in (\mathbb{R}^n)^{\mathcal{K}} \mid h^l(\sigma) < h^l(\bar{\varrho})\right\}.$$

Lastly, if (3), we set

$$S^{\cup} = \left\{h \in (\mathbb{R}^n)^{\mathcal{K}} \mid h^l(\sigma) > h^l(\tau_1)\right\},$$
$$S^{\cap} = \left\{h \in (\mathbb{R}^n)^{\mathcal{K}} \mid h^l(\sigma) < h^l(\tau_2)\right\}.$$

In all cases, it is straightforward to check that $g \notin \mathrm{Fil}_n(\mathcal{K}) \cap \pi^{-1}(\mathrm{block}_L(f|_{\mathcal{K}_m})) \cap S^{\cup} \cap S^{\cap}$, while any $h \in \mathrm{block}_L(f)$ also belongs to $\mathrm{Fil}_n(\mathcal{K}) \cap \pi^{-1}(\mathrm{block}_L(f|_{\mathcal{K}_m})) \cap S^{\cup} \cap S^{\cap}$. Due to the finiteness of both $\mathcal{K}$ and L, we can conclude that $\mathrm{block}_L(f)$ is semiaffine. □

Proof of Lemma 4. Let us fix $f \in \mathrm{Fil}_n(\mathcal{K})$ and denote $\mathrm{MMA}_L(f, \rho_f) = \bigoplus_{i=1}^m \mathbb{I}_i^f$. For any $g \in \mathrm{block}_L(f)$, we have $\mathrm{MMA}_L(g, \rho_g) = \bigoplus_{i=1}^m \mathbb{I}_i^g$, since the vineyard algorithm is determined by the simplex order on each line, which is constant on $\mathrm{block}_L(f)$, and matched families of bars generate corresponding summands. If $i \notin \{1, \ldots, m\}$, then $B_i^f = B_i^g = D_i^f = D_i^g = \emptyset$. Let us now fix $\bar{\imath} \in \{1, \ldots, m\}$ and consider the summands $\mathbb{I}_{\bar{\imath}}^f$, $\mathbb{I}_{\bar{\imath}}^g$ with supports $I_{\bar{\imath}}^f$, $I_{\bar{\imath}}^g$. By construction, these summands are generated by the same family of matched bars across lines. For $h = f, g$, the sets $B_{\bar{\imath}}^h$ and $D_{\bar{\imath}}^h$ are the endpoints of the bars in the $\bar{\imath}$-th matched family (Loiseaux et al., 2025, Rmk. 8). Since the simplex ordering on each line is constant on $\mathrm{block}_L(f)$, the matched families consist of bars generated by the same pairs of simplices, hence of the same number of endpoints. Hence, $\left|B_{\bar{\imath}}^f\right| = |B_{\bar{\imath}}^g|$ and $\left|D_{\bar{\imath}}^f\right| = |D_{\bar{\imath}}^g|$. Since m and these cardinalities are exactly the entries of the configuration, the latter is constant on $\mathrm{block}_L(f)$. □

Proof of Proposition 5. In what follows, we assume the elements of the corner sets are equipped with a canonical ordering (e.g., based on the indexing of the generating simplices and grid lines). This allows us to define $\kappa \colon \mathrm{block}_L(f) \to \mathcal{S}_{\mathcal{C}}$ by

$$\kappa(g) \coloneqq \big((B_1^g, D_1^g), \ldots, (B_m^g, D_m^g)\big),$$

where B_i^g and D_i^g are the birth and death corners of the i-th interval summand of $\mathrm{MMA}_L(g)$. By construction, $e_{\mathcal{C}}(\kappa(g)) = \mathrm{MMA}_L(g)$. It remains to show that κ is semiaffine and locally Lipschitz. For each $\sigma \in \mathcal{K}$ and each $l \in L$, the map $g \mapsto g^l(\sigma)$ is piecewise affine in g, being the maximum of n affine functions. Within $\mathrm{block}_L(f)$, the total preorder is constant, so each bar in $\mathcal{B}(\mathbb{M}(g)|_l)$ is determined by the same pair of simplices $(\sigma_{\mathrm{birth}}, \sigma_{\mathrm{death}})$. Hence the map $g \mapsto (g^l(\sigma_{\mathrm{birth}}), g^l(\sigma_{\mathrm{death}}))$ giving the bar endpoints is piecewise affine. Under the corner computation of Loiseaux et al. (2025, Rmk. 8), each corner is a bar endpoint. Since the simplex pairing is constant within $\mathrm{block}_L(f)$, the same endpoints occur for every $g \in \mathrm{block}_L(f)$, and each corner therefore depends piecewise affinely on g. Since κ collects finitely many such coordinates, κ is piecewise affine on $\mathrm{Fil}_n(\mathcal{K})$, and hence semiaffine on $\mathrm{block}_L(f)$. Finally, κ is continuous since it is a composition of continuous operations (maxima of affine functions). Being piecewise affine and continuous, κ is locally Lipschitz. □

Remark 23. The notion of blocks is similar to, and inspired by, that of cells in o-minimal geometry (Van den Dries, 1998). However, blocks need not be convex. A more restrictive definition (e.g., by requiring that each coordinate induces the same preorder) would yield convex pieces, but Lemma 4 shows that the configuration of MMA is already constant on blocks as defined.

C.2 Proofs of Section 3.2

We first record a criterion for the canonical shift between two interval modules to be a morphism. Let $\mathbb{I}, \mathbb{J}$ be interval modules with supports I, J, and let $\gamma > 0$. Write

$$U = I \cap \mathrm{supp}\, \mathbb{J}(\gamma) = \{x \in I \mid x + \gamma\mathbf{1} \in J\}, \qquad V = J \cap \mathrm{supp}\, \mathbb{I}(\gamma) = \{y \in J \mid y + \gamma\mathbf{1} \in I\},$$

and let $\alpha \colon \mathbb{I} \to \mathbb{J}(\gamma)$ and $\beta \colon \mathbb{J} \to \mathbb{I}(\gamma)$ be the *canonical shifts*, whose components are the identity on U and on V respectively and zero elsewhere. Then α is a morphism if and only if, for every $x \leq_n y$,

(A) $y \in U$ and $x \in I$ imply $x \in U$;

(B) $x \in U$ and $y + \gamma\mathbf{1} \in J$ imply $y \in U$,

and symmetrically for β. This is Proposition 14 of Dey & Xin (2018), stated there in terms of the connected components of U; their intervals are required to be connected, ours are not, but the condition is pointwise and applies to each component separately.

Lemma 24. *Let $\mathbb{I}, \mathbb{J}$ be discretely presented interval modules whose birth corner sets have the same cardinality M, and whose death corner sets have the same cardinality N. Index them as $B^I = \{b_i^I\}_{i\in[M]}$, $B^J = \{b_i^J\}_{i\in[M]}$, $D^I = \{d_j^I\}_{j\in[N]}$, $D^J = \{d_j^J\}_{j\in[N]}$, suppose that corresponding corners have the same infinite coordinates, and that*

$$\|b_i^I - b_i^J\|_\infty \le \gamma \quad \textit{for every } i \in [M], \qquad \|d_j^I - d_j^J\|_\infty \le \gamma \quad \textit{for every } j \in [N].$$

Then $d_I(\mathbb{I}, \mathbb{J}) \le \gamma$.

Proof. Throughout we use Equation (1) in the form: $z \in I$ if and only if $b \le_n z <_n d$ for some $b \in B^I$ and some $d \in D^I$, the two corners being unrelated; and likewise for J.

We verify (A). Let $x \in I$ and $y \in U$ with $x \le_n y$. Since $y + \gamma\mathbf{1} \in J$, there is $d_j^J \in D^J$ with $y + \gamma\mathbf{1} <_n d_j^J$, and since $x \in I$ there is $b_i^I \in B^I$ with $b_i^I \le_n x$. Then

$$b_i^J \ \le_n \ b_i^I + \gamma\mathbf{1} \ \le_n \ x + \gamma\mathbf{1} \ \le_n \ y + \gamma\mathbf{1} \ <_n \ d_j^J,$$

so $x + \gamma\mathbf{1} \in J$ and $x \in U$.

We verify (B). Let $x \in U$, $x \le_n y$ and $y + \gamma\mathbf{1} \in J$. There is $d_j^J \in D^J$ with $y + \gamma\mathbf{1} <_n d_j^J$, and by hypothesis $d_j^J \le_n d_j^I + \gamma\mathbf{1}$, whence $y <_n d_j^I$. From $x \in U \subseteq I$ there is $b_i^I \in B^I$ with $b_i^I \le_n x$, so $b_i^I \le_n x \le_n y <_n d_j^I$ and $y \in I$; together with $y + \gamma\mathbf{1} \in J$ this gives $y \in U$. By the criterion above, α and β are morphisms, the argument for β being the same with the roles of I and J exchanged.

It remains to check that α and β form a γ-interleaving, that is $\beta_{x+\gamma\mathbf{1}} \circ \alpha_x = \varphi_{\mathbb{I}}(x, x + 2\gamma\mathbf{1})$ for every x. Both sides are either the identity or zero. The right-hand side is the identity if and only if $x, x + 2\gamma\mathbf{1} \in I$; the left-hand side if and only if $x \in I$, $x + \gamma\mathbf{1} \in J$ and $x + 2\gamma\mathbf{1} \in I$. It therefore suffices to show that $x, x + 2\gamma\mathbf{1} \in I$ imply $x + \gamma\mathbf{1} \in J$. From $x \in I$ there is $b_i^I \in B^I$ with $b_i^I \le_n x$; from $x + 2\gamma\mathbf{1} \in I$ there is $d_j^I \in D^I$ with $x + 2\gamma\mathbf{1} <_n d_j^I$, that is $x + \gamma\mathbf{1} <_n d_j^I - \gamma\mathbf{1}$. Hence

$$b_i^J \ \le_n \ b_i^I + \gamma\mathbf{1} \ \le_n \ x + \gamma\mathbf{1} \ <_n \ d_j^I - \gamma\mathbf{1} \ \le_n \ d_j^J,$$

so $x + \gamma\mathbf{1} \in J$ and the two sides agree. The verification for β_y composed with $\alpha_{y+\gamma\mathbf{1}}$ is symmetric. □

Recall that $f, g \in \Omega$ by Assumption 6, so that no two distinct simplices take the same value on a line of L: no bar of the L-fibered barcode is empty, and every summand of $\mathrm{MMA}_L(f)$ and of $\mathrm{MMA}_L(g)$ has exactly one birth corner and one death corner per line of L.

Lemma 25. *Let $f \in \mathrm{Fil}_n(\mathcal{K})$, let $g \in \mathrm{block}_L(f)$ and set $\gamma = \|f - g\|_\infty$. Let $\mathbb{I}_i^f$ and $\mathbb{I}_i^g$ be the i-th interval summands of $\mathrm{MMA}_L(f)$ and $\mathrm{MMA}_L(g)$, which correspond to each other under Lemma 4. Each of them has one birth corner and one death corner per line of L, written b_l^f, d_l^f and b_l^g, d_l^g for $l \in L$, and*

$$\|b_l^f - b_l^g\|_\infty \le \gamma, \qquad \|d_l^f - d_l^g\|_\infty \le \gamma.$$

Proof. By Lemma 4 the matching is determined by the orders of the simplices on the lines of L, which are constant on $\mathrm{block}_L(f)$. Hence, on every line $l \in L$, the bar of the i-th summand is generated by the same pair of simplices for f and for g, which we denote by σ and τ. Under the corner computation of Loiseaux et al. (2025, Rmk. 8) the corners are the endpoints of that bar, so

$$b_l^h = c^l + h^l(\sigma)\mathbf{1}, \qquad d_l^h = c^l + h^l(\tau)\mathbf{1}, \qquad h = f, g.$$

The map $h \mapsto h^l(\sigma) = \max_{k\in[n]}(h(\sigma)_k - c_k^l)$ is a maximum of n maps that are 1-Lipschitz for $\|\cdot\|_\infty$, hence is itself 1-Lipschitz, so $|f^l(\sigma) - g^l(\sigma)| \le \|f - g\|_\infty = \gamma$. Since b_l^f and b_l^g share the basepoint c^l,

$$\|b_l^f - b_l^g\|_\infty = \left|f^l(\sigma) - g^l(\sigma)\right| \le \gamma,$$

and the same computation with τ in place of σ gives $\|d_l^f - d_l^g\|_\infty \le \gamma$. □

Lemma 26. *Let $f \in \mathrm{Fil}_n(\mathcal{K})$ and $g \in \mathrm{block}_L(f)$. Then*

$$d_I\big(\mathrm{MMA}_L(f), \mathrm{MMA}_L(g)\big) \le \|f - g\|_\infty.$$

Proof. Set $\gamma = \|f - g\|_\infty$. By Lemma 4 the two descriptors have the same configuration, so their summands are in bijection, and by Lemma 25 corresponding summands have their corners indexed by L and pairwise within γ. Lemma 24 then gives a γ-interleaving between each pair of corresponding summands, and a direct sum of γ-interleavings is a γ-interleaving of the direct sums. □

Proof of Theorem 10. Let $g \in \mathrm{Fil}_n(\mathcal{K})$ with $\|g - f\|_\infty < \varepsilon_f$, let $l \in L$ and let $\varsigma_1 \neq \varsigma_2 \in \mathcal{K}$. Since $h \mapsto h^l(\varsigma)$ is 1-Lipschitz for $\|\cdot\|_\infty$, being a maximum of n maps that are, we have $|g^l(\varsigma) - f^l(\varsigma)| < \varepsilon_f$ for every ς, whence

$$\big|\big(g^l(\varsigma_1) - g^l(\varsigma_2)\big) - \big(f^l(\varsigma_1) - f^l(\varsigma_2)\big)\big| < 2\varepsilon_f \le \big|f^l(\varsigma_1) - f^l(\varsigma_2)\big|.$$

The two differences therefore have the same sign, and in particular the same strict order. As this holds for every line and every pair of simplices, $\le_{g,l} \equiv \le_{f,l}$ for every $l \in L$, that is $g \in \mathrm{block}_L(f)$. The estimate then follows from Lemma 26 applied to $g, g' \in \mathrm{block}_L(f)$, using $\mathrm{block}_L(g) = \mathrm{block}_L(f)$. □

Throughout this subsection f and g are two filtrations lying in adjacent blocks, separated by a single block boundary as in Assumption 8: there is a unique $l^* \in L$ with $\le_{f,l^*} \not\equiv \le_{g,l^*}$, and the two orders differ by a single adjacent transposition of two simplices σ, τ. Recall that under Assumptions 6 and 9 each preorder is a total order and each step involves at most one adjacent transposition. For $c \in \mathbb{R}$ we write $l(c)$ for the diagonal line with basepoint $(c, 0)$. For $h \in \mathrm{Fil}_2(\mathcal{K})$ and distinct $\varsigma_1, \varsigma_2 \in \mathcal{K}$, with $x = h(\varsigma_1)$ and $y = h(\varsigma_2)$, we set $\psi^h_{\varsigma_1\varsigma_2}(c) := h^{l(c)}(\varsigma_1) - h^{l(c)}(\varsigma_2) = \max(x_1 - c, x_2) - \max(y_1 - c, y_2)$, $\zeta_h(\varsigma_1, \varsigma_2) := \max(x_1, y_1) - \max(x_2, y_2)$. We say that ς_1, ς_2 are **strictly comparable** at h if $x <_2 y$ or $y <_2 x$, and **strictly incomparable** at h if neither $x \le_2 y$ nor $y \le_2 x$.

Lemma 27. *Let $h \in \mathrm{Fil}_2(\mathcal{K})$ and let $\varsigma_1 \neq \varsigma_2 \in \mathcal{K}$.*

1. *The map $c \mapsto \psi^h_{\varsigma_1\varsigma_2}(c)$ is monotone on $\mathbb{R}$.*
2. *It takes a strictly positive and a strictly negative value if and only if ς_1, ς_2 are strictly incomparable at h, and in that case it vanishes exactly at $c = \zeta_h(\varsigma_1, \varsigma_2)$, the **exchange point** of the pair.*
3. *$|\zeta_h(\varsigma_1, \varsigma_2) - \zeta_{h'}(\varsigma_1, \varsigma_2)| \le 2\|h - h'\|_\infty$ for all $h, h' \in \mathrm{Fil}_2(\mathcal{K})$.*

Proof. (1) Set $a := x_1 - x_2$ and $b := y_1 - y_2$. Then $\psi^h_{\varsigma_1\varsigma_2}$ is continuous, equals $x_1 - y_1$ for $c \le \min\{a, b\}$ and $x_2 - y_2$ for $c \ge \max\{a, b\}$, and between a and b equals $x_1 - c - y_2$ if $b \le a$ and $x_2 - y_1 + c$ if $a < b$. It is therefore nonincreasing in the first case and nondecreasing in the second.

(2) For $c \le \min(x_1 - x_2, y_1 - y_2)$ we have $\psi(c) = x_1 - y_1$, and for $c \ge \max(x_1 - x_2, y_1 - y_2)$ we have $\psi(c) = x_2 - y_2$; by (1) every value of ψ lies between these two. Hence ψ takes both strict signs if and only if $(x_1 - y_1)(x_2 - y_2) < 0$, that is, if and only if x and y are strictly incomparable. Up to exchanging ς_1 and ς_2, let $x_1 > y_1$ and $x_2 < y_2$. Then $y_1 - y_2 < x_1 - x_2$, and between these two numbers $\psi(c) = x_1 - c - y_2$, which is strictly decreasing and vanishes only at $c = x_1 - y_2 = \max(x_1, y_1) - \max(x_2, y_2)$.

(3) The map $(a, b) \mapsto \max(a, b)$ is 1-Lipschitz for the supremum norm, and ζ_h is a difference of two such maxima of coordinates of h. □

Corollary 28. *Let $h \in \Omega$ and let l, l' be consecutive lines of L.*

1. *Two simplices are ordered differently by $\le_{h,l}$ and $\le_{h,l'}$ if and only if they are strictly incomparable at h and their exchange point lies strictly between c^l_1 and $c^{l'}_1$.*
2. *The orders $\le_{h,l}$ and $\le_{h,l'}$ differ by one adjacent transposition if and only if exactly one pair satisfies (1).*

3. Each pair of simplices changes order at most once along L.

Proof. (1) Since $h \in \Omega$, $\psi^h_{\varsigma_1\varsigma_2}$ vanishes neither at c^l_1 nor at $c^{l'}_1$, and the two orders differ on the pair if and only if these two values have opposite signs; Lemma 27 concludes. (2) If exactly one pair σ, τ changes order, a simplex ς with $\sigma <_{h,l} \varsigma <_{h,l} \tau$ would keep both comparisons on l' and force $\sigma <_{h,l'} \tau$; hence σ, τ are adjacent and the orders differ by their transposition. Conversely, an adjacent transposition changes the order of exactly one pair. (3) Lemma 27(2) gives at most one exchange point. □

Lemma 29. *Let l^- and l^+ be the lines preceding and following l^* in L, and let i be the position occupied by σ and τ in the order on l^*. Exactly one of f and g performs the transposition $(i\ i{+}1)$ on the step $l^- \to l^*$, and the other performs it on the step $l^* \to l^+$; on every other step the two perform the same transposition. If $l^* = l_N$ the same holds with the second step omitted.*

Proof. Write γ^l_h for the permutation corresponding to $\leq_{h,l}$, and set $\pi := \gamma^{l^-}_f = \gamma^{l^-}_g$, the two agreeing by Assumption 8. Let t_f, t_g be the transpositions performed on the step $l^- \to l^*$, so that $\leq_{f,l^*}$ is induced by $t_f\pi$ and $\leq_{g,l^*}$ by $t_g\pi$. By Assumption 9 each of t_f, t_g is either the identity or a single adjacent transposition. Since the two orders on l^* differ by $(i\ i{+}1)$, we get $t_f t_g = (i\ i{+}1)$. This forces one of the two to be the identity: if $t_f = t_g$ their product is the identity, and if they were distinct transpositions the product would be even, while $(i\ i{+}1)$ is odd. Up to exchanging the roles of f and g, we may therefore assume $t_g = \mathrm{id}$ and $t_f = (i\ i+1)$.

If $l^* = l_N$ there is nothing left to prove, so assume l^+ exists and let s_f, s_g be the transpositions performed on the step $l^* \to l^+$. The orders on l^+ agree by Assumption 8, so $s_f t_f \pi$ and $s_g t_g \pi$ induce the same order, giving $s_g = s_f(i\ i{+}1)$. Now $s_f = (i\ i{+}1)$ is impossible: it would make f perform the same transposition twice, whereas by Corollary 28 a pair of simplices changes order at most once along L. And s_f cannot be a different transposition either: $s_g = s_f(i\ i{+}1)$ would then be a product of two distinct transpositions, so an even permutation other than the identity, which Assumption 9 excludes. Hence $s_f = \mathrm{id}$ and $s_g = (i\ i{+}1)$.

Finally, on any step other than these two the orders at both ends agree for f and g, so the transposition performed is the same. □

Proposition 30. *Let $f, g \in \Omega$ lie in adjacent blocks as above, and set $\gamma = \|f - g\|_\infty$. Then*

$$d_I\big(\mathrm{MMA}_L(f), \mathrm{MMA}_L(g)\big) \leq \gamma.$$

Proof. By Lemma 24 it suffices to exhibit a bijection between the summands of $\mathrm{MMA}_L(f)$ and those of $\mathrm{MMA}_L(g)$ under which corresponding corners are within γ. We first produce the bijection, then estimate the corners on the lines $l \neq l^*$, where the two summands share their generators, and finally on l^*, where they need not.

By Lemma 29 we may assume that f performs $(i\ i{+}1)$ on the step $l^- \to l^*$ and nothing on $l^* \to l^+$, that g does the opposite, and that on every other step the two perform the same transposition. Write π for the order on l^-, common to f and g. On every step before l^- the orders at both ends agree for f and g, and the vineyard algorithm of Loiseaux et al. (2025) is deterministic, so the two computations reach l^- with the same basis. The step $l^- \to l^*$ of f has orders π and $(i\ i{+}1)\pi$ at its ends, and so does the step $l^* \to l^+$ of g; both induce the same permutation V and return the same basis, while the two remaining steps perform no transposition and induce the identity. From l^+ on the orders and the bases agree, so the induced permutations coincide. The composite from l^- to l^+ is therefore V for f and V for g; if $l^* = l_N$, there is no line l^+ and the argument stops at l^*. Matching the chains of persistence pairs that carry the same pair on l^- thus gives a bijection under which the two chains carry the same pair on every line $l \neq l^*$, and on l^* carry pairs differing by $(i\ i{+}1)$. By Cohen-Steiner et al. (2006), these pairs differ only in endpoints of the same kind generated by σ and τ, so the bijection preserves the homological degree and the infinite coordinates. This is the required bijection of summands.

Let $\mathbb{I}^f$ and $\mathbb{I}^g$ be matched summands and $l \neq l^*$. If ς generates the birth of their common pair on l, the two birth corners are $c^l + f^l(\varsigma)\mathbf{1}$ and $c^l + g^l(\varsigma)\mathbf{1}$, and since $h \mapsto h^l(\varsigma)$ is 1-Lipschitz,

$$\|b^f_l - b^g_l\|_\infty = \big|f^l(\varsigma) - g^l(\varsigma)\big| \leq \gamma,$$

and likewise for the death corner.

On l^* the same argument applies to matched summands whose pairs agree. Suppose then that they differ, one endpoint being generated by τ for f and by σ for g; by Cohen-Steiner et al. (2006) the exchange is between endpoints of the same kind, so the two corners are both births or both deaths. Since $f, g \in \Omega$ have no ties, the two orders on l^* differ by a strict inversion, say $f^{l^*}(\tau) < f^{l^*}(\sigma)$ and $g^{l^*}(\sigma) < g^{l^*}(\tau)$, and 1-Lipschitzness gives $f^{l^*}(\sigma) \le g^{l^*}(\sigma) + \gamma$ and $g^{l^*}(\tau) \le f^{l^*}(\tau) + \gamma$, whence

$$f^{l^*}(\tau) < f^{l^*}(\sigma) \le g^{l^*}(\sigma) + \gamma, \qquad g^{l^*}(\sigma) < g^{l^*}(\tau) \le f^{l^*}(\tau) + \gamma.$$

Hence $|f^{l^*}(\tau) - g^{l^*}(\sigma)| \le \gamma$, and symmetrically for the other summand involved in the exchange.

Under Assumption 6 no bar is empty, so every summand has one birth and one death corner per line of L. Matched summands therefore have their corners indexed by L and pairwise within γ, and Lemma 24 gives a γ-interleaving between each matched pair. A direct sum of γ-interleavings is a γ-interleaving of the direct sums. □

Proof of Theorem 11. Since $\mathrm{Fil}_n(\mathcal{K})$ is convex, every $h \in [f, g]$ is a valid filtration. Moreover, $[f, g]$ meets finitely many blocks, their number being finite by Proposition 2. There is one boundary $W^l_{\varsigma_1 \varsigma_2}$ for each line of L and each pair of distinct simplices, hence finitely many, and $[f, g]$ meets each of them in finitely many points by Assumption 8. Removing those points leaves finitely many relatively open subsegments, each lying in Ω and, carrying no tie, in a single block. Choose $f = h_0, h_1, \ldots, h_m = g$ in order along $[f, g]$, one in each subsegment.

Two consecutive h_j, h_{j+1} are separated by one point of $[f, g]$, which lies on a single boundary $W^{l^*}_{\sigma\tau}$ with $l^* \neq l_1$ by Assumption 8; by continuity, near that point no other simplex takes a value between those of σ and τ on l^*. Hence h_j and h_{j+1} lie either in the same block, where Lemma 26 applies, or in adjacent blocks whose orders differ only by the adjacent transposition of σ and τ on l^*, where Proposition 30 does; both give

$$d_I\big(\mathrm{MMA}_L(h_j), \mathrm{MMA}_L(h_{j+1})\big) \le \|h_j - h_{j+1}\|_\infty.$$

The h_j are collinear and ordered, so the triangle inequality for d_I gives

$$d_I\big(\mathrm{MMA}_L(f), \mathrm{MMA}_L(g)\big) \le \sum_{j=0}^{m-1} \|h_j - h_{j+1}\|_\infty = \|f - g\|_\infty. \qquad \square$$

C.3 Proofs of Section 3.3

Proof of Theorem 12. Fix $f \in \mathrm{Fil}_n(\mathcal{K})$ and let $\mathcal{C}$ be the configuration of $\mathrm{MMA}_L(f)$. On $\mathrm{block}_L(f)$ Proposition 5 gives $\mathrm{MMA}_L = e_{\mathcal{C}} \circ \kappa$ with κ semiaffine, so $\mathrm{vec} \circ \mathrm{MMA}_L = (\mathrm{vec} \circ e_{\mathcal{C}}) \circ \kappa$ is definable there. The blocks are finitely many and semiaffine by Propositions 2 and 3, so $\mathrm{vec} \circ \mathrm{MMA}_L$ is definable, and so is $\mathcal{L}$, being its composition with the definable maps Φ and E.

Let now $\Phi(x) \in \Omega$ and let $\mathcal{C}$ be the configuration of $\mathrm{MMA}_L(\Phi(x))$. For $h \in \Omega$ we write s_h for the corners of $\mathrm{MMA}_L(h)$, indexed by L. Condition (3) gives $r > 0$ and $\Lambda' \ge 0$ such that $E \circ \mathrm{vec} \circ e_{\mathcal{C}}$ is Λ'-Lipschitz on the $d_{\mathcal{C}}$-ball of radius r around $s_{\Phi(x)}$. For item (i), Theorem 10 puts $B(\Phi(x), \min\{\varepsilon_{\Phi(x)}, r\}) \cap \mathrm{Fil}_n(\mathcal{K})$ inside $\mathrm{block}_L(\Phi(x))$, where Lemma 25 gives $d_{\mathcal{C}}(s_g, s_{g'}) \le \|g - g'\|_\infty$ and $d_{\mathcal{C}}(s_{\Phi(x)}, s_g) < r$. Hence $\mathcal{L}$ is locally Lipschitz at x, since Φ is.

For item (ii), let x' satisfy $\|\Phi(x) - \Phi(x')\|_\infty < r$, and let $\Phi(x) = h_0, \ldots, h_m = \Phi(x')$ be the points chosen in the proof of Theorem 11. Lemma 25 inside a block and the proof of Proposition 30 across a boundary match the summands of $\mathrm{MMA}_L(h_j)$ and $\mathrm{MMA}_L(h_{j+1})$ preserving the homological degree and the infinite coordinates, and move corresponding corners by at most $\|h_j - h_{j+1}\|_\infty$. Every $\mathrm{MMA}_L(h_j)$ therefore has configuration $\mathcal{C}$ and $d_{\mathcal{C}}(s_{h_j}, s_{h_{j+1}}) \le \|h_j - h_{j+1}\|_\infty$, so all s_{h_j} stay within r of s_{h_0} and

$$|\mathcal{L}(x) - \mathcal{L}(x')| \le \Lambda' \sum_{j=0}^{m-1} \|h_j - h_{j+1}\|_\infty = \Lambda' \|\Phi(x) - \Phi(x')\|_\infty. \qquad \square$$

Proof of Corollary 13. Fix $l \in L$ and $\varsigma_1 \neq \varsigma_2 \in \mathcal{K}$. The map $x \mapsto \Phi(x)^l(\varsigma_1) - \Phi(x)^l(\varsigma_2)$ is continuous, since $h \mapsto h^l(\varsigma)$ is 1-Lipschitz, and it never vanishes, since Φ takes values in Ω. As $\mathbb{R}^d$ is connected, its sign is constant. This holds for every line and every pair of simplices, so $\leq_{\Phi(x),l} \equiv \leq_{\Phi(x_0),l}$ for all $x, x_0 \in \mathbb{R}^d$ and $l \in L$, that is $\Phi(\mathbb{R}^d) \subseteq \mathrm{block}_L(\Phi(x_0))$.

Let $\mathcal{C}$ be the configuration of $\mathrm{MMA}_L(\Phi(x_0))$. Proposition 5 gives a semiaffine, locally Lipschitz $\kappa \colon \mathrm{block}_L(\Phi(x_0)) \to \mathcal{S}_\mathcal{C}$ with $\mathrm{MMA}_L = e_\mathcal{C} \circ \kappa$ there, hence

$$\mathcal{L} = (E \circ \mathrm{vec} \circ e_\mathcal{C}) \circ \kappa \circ \Phi \quad \text{on } \mathbb{R}^d.$$

The three factors are definable and locally Lipschitz, and so is their composition. A definable function admits a C^p stratification for every $p \geq 1$, so $\mathcal{L}$ is C^d-stratifiable, and Proposition 18 applies. □

Proposition 31. *Let $\mathcal{C} = (m, p_1, q_1, \ldots, p_m, q_m)$ be a configuration with $n = 2$, and set $P \coloneqq \max_{i \in [m]} p_i q_i$. For $s \in \mathcal{S}_\mathcal{C}$ let $A_i(s)$ be the area of the support of the i-th summand of $e_\mathcal{C}(s)$ cropped to $[0,1]^2$, and let $\widetilde{A}_i(s)$ be as in Equation* (4)*. Then*

$$s \mapsto \alpha \sum_i \widetilde{A}_i - (w + \alpha) \max_i \widetilde{A}_i, \qquad s \mapsto \alpha \sum_i \widetilde{A}_i, \qquad s \mapsto \lambda \sum_i A_i$$

are semialgebraic, and Lipschitz on $(\mathcal{S}_\mathcal{C}, d_\mathcal{C})$. Every term of Equations (5) *and* (6) *is one of these maps.*

Proof. Let β realise $\gamma \coloneqq d_\mathcal{C}(s, s')$ and fix i. By Equation (1) the cropped support of the i-th summand is the union of the $p_i q_i$ cropped rectangles $R_{b_{i,k}, d_{i,k'}} \cap [0,1]^2$, and we pair each with the rectangle of the same indices of the $\beta(i)$-th summand of s'. Clamping to $[0,1]$ is 1-Lipschitz and sends matched infinite coordinates to the same value, so each side of a cropped rectangle moves by at most γ, and the symmetric difference of a pair lies in four strips of width γ and length at most 1. As the symmetric difference of two unions lies in the union of the symmetric differences, $|A_i(s) - A_{\beta(i)}(s')|$ is at most $4 p_i q_i \gamma$. The birth coordinate entering ω is a minimum of matched finite coordinates, hence moves by at most γ, and ω is $1/\Delta$-Lipschitz with values in $[0,1]$, so the adjusted areas of matched summands differ by $O(\gamma)$ as well. Sums and maxima do not depend on the order of the summands, and $|\max_i a_i - \max_i a'_{\beta(i)}| \leq \max_i |a_i - a'_{\beta(i)}|$, which gives the three Lipschitz bounds. Semialgebraicity holds because inclusion and exclusion express A_i through finitely many maxima, minima and products of corner coordinates, and ω is a clamped affine function. □

D Validity of Assumption 9

We keep $n = 2$ and the notation of Appendix C.2.

Remark 32. Let $h \in \Omega$ satisfy the condition of Assumption 9 on consecutive lines l, l' of L. A pair of simplices that keeps its order keeps a strict sign of ψ^h between c_1^l and $c_1^{l'}$ by Lemma 27(1). A pair σ, τ that changes order ties only on $l(\zeta_h(\sigma, \tau))$, where the third clause of Loiseaux et al. (2025, Eq. (22)) orders it as on l. Hence every diagonal line between l and l' carries the permutation of l or of l', which is the hypothesis of Loiseaux et al. (2025, Prop. 16). Genericity enters the proof of that proposition only to make the transposition unique, so the vineyard matching remains compatible under Assumption 9.

For $h_0 \in \mathrm{Fil}_2(\mathcal{K})$ we write $\mathcal{P}(h_0)$ for the set of pairs of distinct simplices that are not strictly comparable at h_0, and set

$$\theta(h_0) \coloneqq \min_{p \neq q \in \mathcal{P}(h_0)} |\zeta_{h_0}(p) - \zeta_{h_0}(q)|, \qquad \nu(h_0) \coloneqq \min_{\{\varsigma_1, \varsigma_2\} \notin \mathcal{P}(h_0)} \min_{k \in \{1,2\}} |h_0(\varsigma_1)_k - h_0(\varsigma_2)_k|,$$

with $\min \emptyset = +\infty$.

Proposition 33. *Let $h_0 \in \mathrm{Fil}_2(\mathcal{K})$ with $\theta(h_0) > \delta$, and set $r \coloneqq \min\{(\theta(h_0) - \delta)/4,\ \nu(h_0)/2\}$. Then every $h \in \Omega$ with $\|h - h_0\|_\infty < r$ satisfies the condition of Assumption 9 on every pair of consecutive lines of L. In particular, Assumption 9 holds for every segment contained in $B(h_0, r)$.*

Proof. A pair outside $\mathcal{P}(h_0)$ is strictly comparable at h_0 with coordinate gaps at least $\nu(h_0)$; each coordinate difference moves by less than $2r \le \nu(h_0)$, so the pair stays strictly comparable at h and, by Corollary 28(1), never changes order along L. Suppose, for a contradiction, that two distinct pairs $p \neq q$ change order between the same consecutive lines l, l' at h. By Corollary 28(1) both are strictly incomparable at h, hence lie in $\mathcal{P}(h_0)$, and both exchange points lie strictly between c_1^l and $c_1^{l'}$, which are δ apart, so $|\zeta_h(p) - \zeta_h(q)| < \delta$. On the other hand, Lemma 27(3) gives

$$|\zeta_h(p) - \zeta_h(q)| \ge |\zeta_{h_0}(p) - \zeta_{h_0}(q)| - 4\|h - h_0\|_\infty > \theta(h_0) - 4r \ge \delta.$$

Hence at most one pair changes order between any two consecutive lines, and Corollary 28(2) concludes. The last claim holds because balls are convex. □

For almost every $h_0 \in \mathrm{Fil}_2(\mathcal{K})$ we have $\theta(h_0) > 0$, so Proposition 33 applies around h_0 for every L with spacing $\delta < \theta(h_0)$. Indeed, a pair of $\mathcal{P}(h_0)$ that is not strictly incomparable places h_0 on a hyperplane $\{h(\varsigma_1)_k = h(\varsigma_2)_k\}$; otherwise $\zeta_{h_0}(p) = h_0(\varsigma)_1 - h_0(\varsigma')_2$ for distinct simplices $\varsigma \neq \varsigma'$ of p, and $\zeta_{h_0}(p) = \zeta_{h_0}(q)$ is a nontrivial affine equation for $p \neq q$, so $\theta(h_0) = 0$ only on finitely many hyperplanes.

E CVAE ARCHITECTURE AND TRAINING DETAILS

E.1 CONDITIONING AND ARCHITECTURE

The conditioning map c contains two channels of Gaussian bumps centered on the prescribed objects, merged by a pointwise maximum within each channel: channel 0 represents filled objects and channel 1 represents hollow objects. The bump of object j has standard deviation $\sigma_j = \max\{R_j/2,\, 1\}$ pixels, where R_j is the object radius. These maps therefore provide object locations, topology, and scales, but not the shape of an object. The spatial distributions used by the topological loss are supplied separately (Appendix E.4).

Encoder. The encoder concatenates the target mask x and conditioning map c along the channel dimension. Four convolutions with kernel size 4, stride 2, and padding 1 produce 64, 128, 256, and 512 channels, respectively. Each convolution is followed by a LeakyReLU activation with negative slope 0.2; the last three additionally use GroupNorm with 32 groups before the activation. Two linear heads predict the mean and log-variance of the diagonal Gaussian posterior over $z \in \mathbb{R}^{d_z}$, with $d_z = 32$:

$$q_\phi(z \mid x, c) = \mathcal{N}\big(\mu_\phi(x, c), \mathrm{diag}\big(\sigma_\phi^2(x, c)\big)\big)\,. \tag{9}$$

Latent samples are obtained through reparameterization:

$$z = \mu_\phi(x, c) + \sigma_\phi(x, c) \odot \epsilon, \qquad \epsilon \sim \mathcal{N}(0, I). \tag{10}$$

Decoder. For square inputs of side H divisible by 16 (here $H = 64$), a linear layer projects z into a $512 \times (H/16) \times (H/16)$ feature map. Four upsampling blocks produce 256, 128, 64, and 32 output channels. Each block applies SPADE, LeakyReLU with negative slope 0.2, bilinear upsampling by a factor of two, and a 3×3 convolution, followed by a second SPADE–LeakyReLU–convolution sequence. The decoder convolutions use replication padding.

For a feature map a_ℓ, spatially-adaptive normalization is

$$\mathrm{SPADE}_\ell(a_\ell, c) = \big(1 + \gamma_\ell(c)\big) \odot \mathrm{GN}(a_\ell) + \eta_\ell(c), \tag{11}$$

where GN is GroupNorm with 32 groups and without learnable affine parameters. The conditioning map is bilinearly resized to the feature resolution and passed through a shared 3×3 convolution with 128 channels and ReLU, followed by separate 3×3 convolutions producing γ_ℓ and η_ℓ. A final 3×3 convolution produces logits $\ell_\theta(z, c)$, with foreground probabilities $\hat{x} = \mathrm{sigmoid}(\ell_\theta(z, c))$. At inference, the decoder receives $z \sim \mathcal{N}(0, I)$ together with c.

E.2 RECONSTRUCTION AND LATENT REGULARIZATION

For a training pair (x, c), the CVAE objective is

$$\mathcal{L}_{\mathrm{CVAE}}(x, c) = \mathbb{E}_{z \sim q_\phi(z \mid x, c)}\left[\mathcal{L}_{\mathrm{BCE}}(x, \hat{x})\right] + \beta_t \mathcal{L}_{\mathrm{KL}}, \tag{12}$$

where reconstruction is measured over the image vertices V:

$$\mathcal{L}_{\mathrm{BCE}}(x, \hat{x}) = -\sum_{v \in V} \left[x(v) \log \hat{x}(v) + (1 - x(v)) \log(1 - \hat{x}(v)) \right]. \tag{13}$$

In implementation, this expression is evaluated directly from the logits for numerical stability, and a total-variation penalty $\lambda_{\mathrm{TV}}\,\mathrm{TV}(\hat{x})$ is added to it, where TV is the sum of the mean absolute differences between horizontally and between vertically adjacent pixels. The divergence from the standard Gaussian prior is

$$\mathcal{L}_{\mathrm{KL}} = D_{\mathrm{KL}}\left(q_\phi(z \mid x, c) \,\|\, \mathcal{N}(0, I)\right) = \frac{1}{2} \sum_{k=1}^{d_z} \left(\mu_k^2 + \sigma_k^2 - \log \sigma_k^2 - 1 \right). \tag{14}$$

The coefficient β_t combines a linear warmup with a controller that keeps the KL divergence close to a target value KL^*:

$$\beta_t = \tilde{\beta}_t \min\left\{1, \frac{t}{T_{\mathrm{KL}}}\right\}, \tag{15}$$

where $\tilde{\beta}_t = \beta_0$ for $t < T_{\mathrm{KL}}$ and, for $t \geq T_{\mathrm{KL}}$,

$$\log \tilde{\beta}_{t+1} = \mathrm{clip}\left(\log \tilde{\beta}_t + \eta\, \frac{\mathrm{KL}_t - \mathrm{KL}^*}{\mathrm{KL}^*},\ \log 10^{-5},\ \log 10^{3} \right), \tag{16}$$

with KL_t the minibatch average of $\mathcal{L}_{\mathrm{KL}}$ at iteration t.

E.3 Loss estimation and schedules

Equation (8) contains two topological terms, one on posterior reconstructions and one on prior samples. The posterior expectation is estimated using one reparameterized sample per image, shared by the reconstruction and topological terms. The prior expectation uses one sample $z \sim \mathcal{N}(0, I)$ per image, decoded with the same conditioning map, so the prior term updates only the decoder. The two topological terms alternate: the posterior term is evaluated at even iterations and the prior term at odd iterations. At each evaluation, persistence is computed for $k = 4$ of the $B = 16$ images of the shuffled minibatch, selected as a cyclically shifted block of consecutive indices. Reconstruction and KL losses are averaged over the minibatch, while topological terms are averaged over the k selected images.

The topological terms are activated at iteration t_0. Thereafter, the coefficient penalizing additional features is increased according to

$$\alpha_t = \alpha \min\left\{1, \frac{t - t_0}{T_{\mathrm{warm}}}\right\}, \qquad t \geq t_0, \tag{17}$$

where $T_{\mathrm{warm}} > 0$ is the warmup duration. The reward coefficients w_0, w_1 and the outside coefficient λ_{out} remain fixed. Training uses Adam; Table 2 lists all hyperparameters.

E.4 Topological-loss conventions and regularity

Spatial distributions. For object j with center p_j and radius R_j, the spatial distribution is the Lorentzian (Cauchy) profile

$$g_j(p) = \left(1 + \frac{\|p - p_j\|^2}{\gamma_j^2}\right)^{-1}, \qquad \gamma_j = \sqrt{3/2}\, R_j, \tag{18}$$

for filled and hollow objects alike. Whether a hole is required enters only through the label h_j in Equation (5). The complete distribution is $g_{\mathrm{mix}} = \max_j g_j$. Each g_j, as well as g_{mix}, is min–max normalized to $[0, 1]$ on the image grid.

Bifiltrations. The 64×64 pixel grid is the vertex set of a cubical complex $\mathcal{K}$ whose edges and squares join adjacent pixels. Each bifiltration is defined on the pixels and extended to higher cells by the coordinatewise maximum over their vertices. The arguments of Section 3 involve only the cells of $\mathcal{K}$ and their face relations, and apply to this cubical complex. For each object j, MMA is evaluated separately on $f_j = (1 - \hat{x}, 1 - g_j)$ in homological degrees zero and one, using $|L| = 25$ lines in a box computed once from a ground-truth bifiltration and kept fixed during training.

Table 2: Training hyperparameters of the MMA-supervised CVAE.

Group	Parameter	Value
Optimization	Learning rate	$3 \cdot 10^{-5}$
	Batch size B, epochs	16, 120
Model	Image size, latent dimension d_z	64×64, 32
Reconstruction	λ_{TV}	$5 \cdot 10^{-4}$
KL	β_0, T_{KL}	10^{-3}, 500
	KL^*, η	60 nats, 10^{-2}
Topological schedule	λ_{topo}, λ_{gen}	1.5, 0.3
	t_0, T_{warm}	200, 200
	Images per evaluation k	4
MMA	Number of lines $\lvert L\rvert$	25
	Resolution	64×64
Individual loss	w_0, w_1, α	10, 15, 5
	τ, Δ	0.9, 0.1
Outside loss	λ_{out}	2

Summand weights. The support area $A_{j,i}^{H_q}$ of each summand is cropped to $[0,1]^2$ and weighted as in Equation (4), where $b_{j,i}^{H_q}$ is the smallest finite first coordinate among the summand's birth corners. If no such coordinate exists, its weight is set to one. The ramp in Equation (3) equals one for $b \le \tau - \Delta$, decreases linearly on $[\tau - \Delta, \tau]$, and vanishes for $b \ge \tau$. It is $1/\Delta$-Lipschitz. All summands are retained without hard size thresholds or truncation.

Individual and outside terms. Equation (5) rewards the largest adjusted degree-zero area with coefficient $-w_0$. For hollow objects, it also rewards the largest adjusted degree-one area with coefficient $-w_1$. Every remaining adjusted area is penalized with coefficient α; for filled objects, this includes all degree-one areas. The largest adjusted area is defined as zero when the corresponding family is empty.

For outside supervision, the spatial coordinate is g_{mix} rather than $1 - g_{\mathrm{mix}}$, so low-density regions enter early. All degree-zero support areas of $f_{\mathrm{out}} = (1-\hat{x}, g_{\mathrm{mix}})$ are penalized without birth weighting. This term is computed once per image, independently of the number of prescribed objects.

Regularity. The map from the network parameters to each bifiltration is a composition of affine maps, LeakyReLU, GroupNorm, bilinear upsampling, the sigmoid, and coordinatewise maxima; it is therefore locally Lipschitz and definable in the o-minimal structure $\mathbb{R}_{\mathrm{exp}}$, which gives condition (1) of Theorem 12. Each topological contribution is a functional of the MMA of one bifiltration and, taking vec to be the contribution itself and E the identity, satisfies conditions (2) and (3) by Proposition 31. Consequently, the theorem applies to each contribution, and its conclusions extend to their finite sum at parameters where all constituent bifiltrations lie in Ω.

F Evaluation metric

This appendix defines the score used in Section 4.2. All distances are in pixels of the 64×64 grid. For an image with N targets, where target j has center p_j, radius R_j and prescribed degree $h_j \in \{0, 1\}$, the score is

$$S = \frac{1}{N}\Big(N_{\mathrm{spur}} + \sum_{j=1}^{N} \chi_j\Big), \qquad \chi_j = \begin{cases} 1, & \text{unmatched,} \\ 0.5, & \text{matched, degree } 1 - h_j, \\ \frac{1}{4}\big(\min\{u_j/R_j, 1\} + v_j\big), & \text{matched, degree } h_j, \end{cases} \tag{19}$$

where N_{spur} is the number of spurious features, u_j is the distance from the matched feature to p_j, and $v_j \in [0, 0.95]$ is the quality deficit defined below. Features, matching and v_j are defined in the following paragraphs.

Detection. The generated mask $\hat{x}$ is thresholded at 0.5. Foreground pixels are connected through their eight neighbours and background pixels through their four neighbours, so that a ring whose pixels touch only diagonally still encloses its hole. A hole is a background component that does not touch the image border; each hole is a feature of degree 1, located at its centroid. Each foreground component that encloses no hole and contains at least 5 pixels is a feature of degree 0, located at its centroid; smaller components are discarded as noise. A component enclosing several holes therefore contributes one degree-one feature per hole and no degree-zero feature; in particular, a filled object that touches a hollow one yields no feature of its own.

Matching. All feature–target pairs are sorted by the distance between the feature position and the target center p_j, and accepted greedily in increasing order when neither element is already assigned and the distance is at most $R_j + 5$. The matching ignores degrees, which are compared only in Equation (19). Targets left unassigned are unmatched, and features left unassigned are spurious.

Quality deficit. The deficit v_j is computed only for a target matched to a feature of the prescribed degree h_j. Let W_j be the square window of half-side $\mathrm{round}(R_j)+5$ centered at the feature position and clipped to the image. On W_j we compute the MMA of the bifiltration $(1-\hat{x},\, 1-g_{\mathrm{mix}})$, where $\hat{x}$ is the unthresholded output and g_{mix} is the complete distribution of Appendix E.4, normalized over the whole image before cropping. The bifiltration is extended by coordinatewise maxima to the triangulation of W_j obtained by splitting every pixel square along its main diagonal. We use the default settings of multipers (557 lines) and cap death coordinates at 1. Each summand $\mathbb{I}$ with birth corners B and death corners D receives the weight

$$w(\mathbb{I}) = \frac{1}{2} \max_{b\in B,\, d\in D} \min_{k\in\{1,2\}} (d_k - b_k)^+ = d_I(\mathbb{I}, 0) \in \left[0, \tfrac{1}{2}\right], \tag{20}$$

and the dominant summand of degree h_j is the one of largest weight w_j. If no summand of degree h_j exists or $w_j < 0.05$, we set $v_j = 0$, since detection has already confirmed the degree. Otherwise

$$v_j = \max\left\{(1 - 2w_j - 0.05)^+,\; h_j\,(1 - \delta_j - 0.05)^+\right\}, \tag{21}$$

where δ_j is the largest second coordinate among the death corners of the dominant summand. The first term measures how far the dominant summand is from the maximal weight $1/2$, attained by a feature born at the origin that survives the whole unit box. The second, used for hollow objects only, penalizes a hole that dies before the spatial filtration is exhausted. The offset 0.05 absorbs discretization jitter and gives $v_j \in [0, 0.95]$.

Aggregation. In the last case of Equation (19), $\chi_j \le \frac{1}{4}(1 + 0.95) = 0.4875$, so a missing or spurious object costs more than a wrong degree, which costs more than any error in position or quality. Object size and shape are not scored. Because spurious features are also normalized by N, S can exceed 1. For each object count we report the mean and standard deviation of S over the 100 test images; the same procedure and constants are used for every model.