Functorial Manifold Learning

Anonymous authors
Paper under double-blind review

Abstract

We adapt previous research on category theory and topological unsupervised learning to develop a functorial perspective on manifold learning. We first characterize manifold learning algorithms as functors that map pseudometric spaces to optimization objectives and factor through hierarchical clustering functors. We then use this characterization to prove refinement bounds on manifold learning loss functions and construct a hierarchy of manifold learning algorithms based on their invariants. We express several popular manifold learning algorithms as functors at different levels of this hierarchy and present bounds on how closely the embeddings these algorithms produce from noisy data approximate the embeddings they would learn from noiseless data.

1 Manifold Learning

A manifold learning algorithm constructs an $n \times m$ real valued matrix of embeddings in $\mathbf{Mat}_{n,m} = \mathbb{R}^{n*m}$ from a finite pseudometric space with n points. In this work we focus on algorithms that operate by solving **embedding optimization problems**, or tuples (n, m, l) where $l : \mathbf{Mat}_{n,m} \to \mathbb{R}$ is a loss function. We call the set of all $A \in \mathbf{Mat}_{n,m}$ that minimize $l(A)$ the **solution set** of the embedding optimization problem.

Formally we define a **manifold learning problem** to be a function that maps the pseudometric space (X, d_X) to an embedding optimization problem of the form $(|X|, m, l)$. Note that this definition does not specify how the optimization problem is solved. For example, the Metric Multidimensional Scaling manifold learning problem maps (X, d_X) to $(|X|, m, l)$ where $l(A) = (d_X(x_i, x_j) - ||A_i - A_j||)^2$. Optimizing this objective involves finding a matrix A such that the Euclidean distance matrix of the rows of A is as close as possible to the d_X distance matrix.

If a manifold learning problem maps isometric pseudometric spaces to embedding optimization problems with the same solution set, we call it **isometry-invariant**. Intuitively, isometry-invariant manifold learning algorithms do not change their output based the ordering of X. One property of these problems is that they factor through hierarchical clustering:

Proposition 1. *Given any isometry-invariant manifold learning problem M there exists a manifold learning problem $L \circ H$, where H is a hierarchical overlapping clustering algorithm (as defined by Shiebler (2020)) and L is a function that maps the output of H to an embedding optimization problem, such that the solution spaces of the images of M and $L \circ H$ on any pseudometric space (X, d_X) are identical. (Proof in Appendix B.1)*

Intuitively, Proposition 1 holds because manifold learning problems generate loss functions by grouping points in the finite pseudometric space together. In order to use this property to adapt clustering theorems into manifold learning theorems we will introduce a target category of optimization problems and replace functions with functors:

Definition 1. *The objects in $\mathbf{Loss}_m$ are tuples $(n, \{c_{ij}, e_{ij}\})$ where n is a natural number, $c_{ij}, e_{ij} : \mathbb{R}_{\geq 0} \to \mathbb{R}$ are real-valued functions that satisfy $c_{i'j'}(x) = e_{i'j'}(x) = 0$ for $i' > n$ or $j' > n$. $\mathbf{Loss}_m$ is a preorder where $(n, \{c_{ij}, e_{ij}\}) \leq (n', \{c'_{ij}, e'_{ij}\})$ iff for any $x \in \mathbb{R}_{\geq 0}, i, j \in \mathbb{N}$ we have $c'_{ij}(x) \leq c_{ij}(x)$ and $e_{ij}(x) \leq e'_{ij}(x)$.*

Definition 2. *The objects in the category $\mathbf{FLoss}_m$ are functors $F : (0, 1]^{op} \to \mathbf{Loss}_m$ that commute with the forgetful functor that maps $(n, \{c_{ij}, e_{ij}\})$ to n. The morphisms in $\mathbf{FLoss}_m$ are natural transformations. We call n the **cardinality** of F.*

Each $F \in \mathbf{FLoss}_m$ corresponds to an embedding optimization problem. We can write $F(a) = (n, \{c_{F(a)_{ij}}, e_{F(a)_{ij}}\})$ and define the embedding optimization problem $(n, m, \mathbf{l}_F)$ where the F-loss $\mathbf{l}_F(A)$

is $\mathbf{l}_F(A) = \int_{a\in(0,1]} \sum_{\substack{i\in 1...n\\ j\in 1...n}} c_{F(a)_{ij}}(\|A_i - A_j\|) + e_{F(a)_{ij}}(\|A_i - A_j\|)\, da$. Intuitively, $\sum_{\substack{i\in 1...n\\ j\in 1...n}} c_{F(a)_{ij}}(\|A_i - A_j\|) + e_{F(a)_{ij}}(\|A_i - A_j\|)$ is a loss term that exists with the strength a, and $\mathbf{l}_F(A)$ is the average loss across all strengths.

Definition 3. *Suppose* **PMet** *is the category of pseudometric spaces and non-expansive maps and* **FCov** *is the category of fuzzy flag covers and natural transformations (Culbertson et al., 2016; Shiebler, 2020). Then given the subcategories* $\mathbf{D} \subseteq \mathbf{PMet}, \mathbf{D}' \subseteq \mathbf{FCov}$, *the composition* $L \circ H : \mathbf{D} \to \mathbf{FLoss}_m$ *forms a* **D*-manifold learning functor*** *if* $H : \mathbf{D} \to \mathbf{D}'$ *is a hierarchical* **D***-clustering functor and* $L : \mathbf{D}' \to \mathbf{Loss}_m$ *is a functor that maps a fuzzy flag cover with vertex set* X *to some* $F_X \in \mathbf{FLoss}_m$ *with cardinality* $|X|$.

Intuitively a manifold learning functor $\mathbf{D} \xrightarrow{H} \mathbf{D}' \xrightarrow{L} \mathbf{FLoss}_m$ factors through a hierarchical clustering functor and sends (X, d_X) to F where $F(a) = (|X|, \{c_{F(a)_{ij}}, e_{F(a)_{ij}}\})$. We will say that $M = L \circ H$ is **standard form** if M maps the one-point metric space $(\{*\}, 0)$ to some F where $c_{F(a)_{ij}}(x) = e_{F(a)_{ij}}(x) = 0$ and $\forall \epsilon, \delta \in \mathbb{R}_{\geq 0}, H(X, d_X + \epsilon)(-log(\delta)) \simeq H(X, d_X)(-log(\delta + \epsilon))$. Each manifold learning functor corresponds to a manifold learning problem that maps (X, d_X) to $(|X|, m, \mathbf{l}_{M(X,d_X)})$.

1.1 A Spectrum of Manifold Learning Functors

Recall the single and maximal linkage hierarchical overlapping clustering algorithms $\mathcal{SL}$ and $\mathcal{ML}$ which map the pseudometric space (X, d_X) to the fuzzy non-nested cover (X, C_{X_a}) where C_{X_a} is respectively the connected components of the $-log(a)$-Vietoris-Rips complex of (X, d_X) and the maximally linked sets of the relation R_a in which $x_1 R_a x_2$ if $d_X(x_1, x_2) \leq -log(a)$ (Shiebler, 2020; Culbertson et al., 2016). If we apply functoriality to Proposition 6 in Shiebler (2020) we see:

Proposition 2. *Suppose* $\mathbf{D}$ *is a subcategory of* **PMet** *such that* $\mathbf{PMet}_{bij} \subseteq \mathbf{D}$, $L \circ H$ *is a* **D***-manifold learning functor such that* H *is non-trivial and for all* $a \in (0, 1]$, *the functor* $H(-)(a) : \mathbf{D} \to \mathbf{Cov}$ *has clustering parameter* $\delta_{H,a}$. *Then for* $a \in (0, 1]$ *and* $(X, d_X) \in \mathbf{D}$ *we have maps:*

$$(L \circ \mathcal{ML})(X, d_X)(e^{-\delta_{H,a}}) \leq (L \circ H)(X, d_X)(a) \leq (L \circ \mathcal{SL})(X, d_X)(e^{-\delta_{H,a}}) \tag{1}$$

That are natural in a *and* (X, d_X). *(Proof in Appendix B.2)*

Intuitively, this proposition states that every manifold learning functor maps (X, d_X) to a loss that is both no more interconnected than the loss that does not distinguish points within the same connected component of the Vietoris-Rips complex and no less interconnected than the loss that treats each pair of points independently. There are many manifold learning functors that lie between these extremes. In particular, for any functor $L : \mathbf{PMet}_{inj} \to \mathbf{Loss}_m$ and sequence of clustering functors $\mathcal{ML}, H_1, H_2, ..., H_n, \mathcal{SL}$ whose outputs refine each other we can apply functoriality to form a sequence of manifold learning functors $L \circ \mathcal{ML} \leq L \circ H_1 \leq ... \leq L \circ H_n \leq L \circ \mathcal{SL}$. For example, consider the $\mathcal{L}_k$ family of hierarchical overlapping clustering functors from Culbertson et al. (2016): for $k \in \mathbb{N}$, the cover $\mathcal{L}_k(X, d_X)(a)$ is the maximal linked sets of the relation R_a where $xR_a x'$ if there is a sequence $x = x_1, x_2..., x_{k-1}, x_k = x'$ in X where $d(x_i, x_{i+1}) \leq -log(a)$. The functor $L \circ \mathcal{L}_k$ therefore maps (X, d_X) to a loss that distinguishes only between points whose shortest path in the Vietoris-Rips complex is longer than k. For $k > 1$ this loss is more interconnected than $L \circ \mathcal{ML}$ and less interconnected than $L \circ \mathcal{SL}$. This also suggests a recipe for generating new manifold learning algorithms (see Appendix A): first express an existing manifold learning problem in the form $L \circ H$, and then form $L \circ \mathcal{SL}, L \circ \mathcal{ML}$, or any of the functors along the spectrum $L \circ \mathcal{L}_k$.

1.2 Characterizing Manifold Learning Problems

Similarly to how Carlsson & Mémoli (2013) characterize clustering algorithms in terms of their functoriality over different subcategories of pseudometric spaces, we can characterize manifold learning algorithms based on the subcategory $\mathbf{D} \subseteq \mathbf{PMet}$ over which they are functorial.

We have already introduced the class of isometry-invariant manifold learning problems. Any $\mathbf{PMet}_{isom}$-manifold learning functor is isometry-invariant, and an isometry-invariant manifold learning problem is **expansive-contractive** if the loss that it aims to minimize decomposes into the sum of an expansive term e that decreases as distances increase and a contractive term c that increases as distances increase. Intuitively, expansive-contractive manifold learning problems use the term e to push together points that are close in the original space and use the term c to push apart points that are far in the original space. Any $\mathbf{PMet}_{bij}$-manifold learning functor is expansive-contractive.

An expansive-contractive manifold learning problems is **positive extensible** if c increases and e decreases when we increase $|X|$. If instead c decreases and e increases when we increase $|X|$, we say it is **negative extensible**. Any $\mathbf{PMet}_{sur}$-manifold learning functor is positive extensible and any $\mathbf{PMet}_{inj}$-manifold learning functor is negative extensible.

1.2.1 Metric Multidimensional Scaling ($\mathbf{PMet}_{sur}$-Manifold Learning Functor)

The most straightforward strategy for learning embeddings is to minimize the difference between the pairwise distance matrix of the original space and the pairwise Euclidean distance matrix of the learned embeddings. The **Metric Multidimensional Scaling** algorithm (Abdi, 2007) does exactly this. Given a finite pseudometric space (X, d_X), the Metric Multidimensional Scaling embedding optimization problem is $(|X|, m, l)$ where $l(A) = \sum_{\substack{i\in 1...n \\ j\in 1...n}} (d_X(x_i, x_j) - \|A_i - A_j\|)^2$. When the distance matrix of the pseudometric space is double-centered (mean values of rows/columns are zero) this is the same objective that Principal Components Analysis (PCA) optimizes (Hinton, 2013).

Proposition 3. *There exists a functor $MDS : \mathbf{FCov}_{sur} \to \mathbf{FLoss}_m$ such that the $\mathbf{PMet}_{sur}$-manifold learning functor $MDS \circ \mathcal{ML}$ maps the finite pseudometric space (X, d_X) to the Metric Multidimensional Scaling embedding optimization problem. (Proof in Appendix B.3)*

1.2.2 IsoMap ($\mathbf{PMet}_{sur}$-Manifold Learning Functor)

For many real world datasets it is the case that the distances between nearby points are more reliable and less noisy than the distances between far away points. The **IsoMap** algorithm (Tenenbaum et al., 2000) redefines the distances between far apart points in terms of the distances between near points. Given a finite pseudometric space (X, d_X), the IsoMap embedding optimization problem is $(|X|, m, l)$ where $l(A) = \sum_{\substack{i\in 1...n \\ j\in 1...n}} (d'_X(x_i, x_j) - \|A_i - A_j\|)^2$ such that $d'_X(x_i, x_j)$ is the length of the shortest path between x_i and x_j in the graph in which there exists an edge of length $d_X(x, x')$ between each pair of points $(x, x') \in X$ with $d_X(x, x') \leq \delta$.

Proposition 4. *For any $\delta \in \mathbb{R}_{\geq 0}$, there exists a hierarchical $\mathbf{PMet}$-clustering functor $IsoCluster_\delta$ such that the $\mathbf{PMet}_{sur}$-manifold learning functor $MDS \circ IsoCluster_\delta$ maps the finite pseudometric space (X, d_X) to the IsoMap embedding optimization problem. (Proof in Appendix B.4)*

1.2.3 UMAP ($\mathbf{PMet}_{isom}$-Manifold Learning Functor)

The UMAP algorithm builds a local uber-metric space around each point in X, converts each local uber-metric space to a fuzzy simplicial complex, and minimizes a loss function based on a fuzzy union of these fuzzy simplicial complexes. Given a finite pseudometric space (X, d_X), the UMAP embedding optimization problem is $(|X|, m, l)$ where l is the fuzzy cross-entropy:

$$l(A) = \sum_{\substack{i\in 1...|X| \\ j\in 1...|X|}} W_{ij}\, log\left(\frac{W_{ij}}{e^{-\|A_i - A_j\|}}\right) + (1 - W_{ij})\, log\left(\frac{1 - W_{ij}}{1 - e^{-\|A_i - A_j\|}}\right)$$

and W_{ij} is the weight of the fuzzy union of the 1-simplices connecting x_i and x_j in the Vietoris-Rips complexes formed from the $|X|$ local uber-metric spaces (X, d_{x_i}) where:

$$d_{x_i}(x_j, x_k) = \begin{cases} d_X(x_j, x_k) - min_{l=1...n} d_X(x_i, x_l) & i = j, i = k \\ \infty & else \end{cases}$$

Proposition 5. *There exists a hierarchical $\mathbf{PMet}_{isom}$-clustering functor FuzzySimplex that decomposes into the composition of four functors that: build a local uber-metric space around each point in X, convert each local uber-metric space to a fuzzy simplicial complex, take a fuzzy union of these fuzzy simplicial complexes, and convert the resulting fuzzy simplicial complex to a fuzzy non-nested flag cover. (Proof in Appendix B.5)*

Proposition 6. *There exists a functor $FCE : \mathbf{FCov}_{bij} \to \mathbf{FLoss}_m$ such that the composition $FCE \circ$ FuzzySimplex is a $\mathbf{PMet}_{isom}$-manifold learning functor that maps the pseudometric space (X, d_X) to the UMAP embedding optimization problem. (Proof in Appendix B.6)*

Since the UMAP distance rescaling procedure does not preserve non-expansive maps, even if a map from (X, d_X) to $(X', d_{X'})$ is non-expansive, this will not necessarily be the case for all of the local uber-metric spaces (X, d_{x_i}) that we build from (X, d_X) and $(X', d_{X'})$. For this reason $FCE \circ FuzzySimplex$ is not functorial over $\mathbf{PMet}_{bij}$.

2 Stability of Manifold Learning Algorithms

We can use this functorial perspective on manifold learning to reason about the stability of manifold learning algorithms under dataset noise. An **ϵ-interleaving** between the functors $F, G : \mathbb{R}_{\geq 0} \to \mathbf{C}$ is a collection of commuting natural transformations between $F(\delta) \to G(\delta + \epsilon)$ and $G(\delta) \to F(\delta + \epsilon)$ (Chazal et al., 2014; Scoccola, 2020). The **interleaving distance** d_I between such functors is the smallest ϵ such that an ϵ-interleaving exists. In order to study interleavings between functors in $\mathbf{FCov}$ or $\mathbf{FLoss}_m$ whose domain is $(0, 1]^{op}$ rather than $\mathbb{R}_{\geq 0}$, we will say that the functors F, G are ϵ_*-interleaved when there is an ϵ-interleaving between the functors $F \circ r$ and $G \circ r$ where $r(x) = e^{-x}$. We will also write $d_{I_*}(F, G) = d_I(F \circ r, G \circ r)$.

Proposition 7. *Given a subcategory* $\mathbf{D}$ *of* $\mathbf{PMet}$*, a standard form* $\mathbf{D}$*-manifold learning functor* $M = L \circ H$ *and a pair of finite pseudometric spaces* $(X, d_X), (Y, d_Y)$ *such that there exists a pair of morphisms* $f : (X, d_X) \to (Y, d_Y + \epsilon), g : (Y, d_Y) \to (X, d_X + \epsilon)$ *in* $\mathbf{D}$*, we have* $d_{I_*}(M(X, d_X), M(Y, d_Y)) \leq \epsilon$*. (Proof in Appendix B.7)*

Proposition 7 is similar in spirit to previous results that use the Gromov-Hausdorff distance between metric spaces to bound the distance between their corresponding Rips complexes (Chazal et al., 2014; Bubenik & Scott, 2014; Scoccola, 2020; Blumberg & Lesnick, 2017). As a special case, if M is an $\mathbf{PMet}_{bij}$-manifold learning functor and there exists an ϵ-isometry between $(X, d_X), (Y, d_Y)$ then $d_{I_*}(M(X, d_X), M(Y, d_Y)) \leq \epsilon$. We can use this to show:

Proposition 8. *Suppose we have a standard form* $\mathbf{PMet}_{sur}$*-manifold learning functor* M*, a pair of* ϵ*-isometric finite pseudometric spaces* $(X, d_X), (Y, d_Y)$ *and the matrices* A_X, A_Y *that respectively minimize* $\mathbf{l}_{M(X,d_X)}$ *and* $\mathbf{l}_{M(Y,d_Y)}$*. Then if* $|c_{M(X,d_X)(a)_{ij}}(x)| \leq \frac{K_\mathbf{c}}{2}, |c_{M(Y,d_Y)(a)_{ij}}(x)| \leq \frac{K_\mathbf{c}}{2}$ *and* $|e_{M(X,d_X)(a)_{ij}}(x)| \leq \frac{K_\mathbf{e}}{2}, |e_{M(Y,d_Y)(a)_{ij}}(x)| \leq \frac{K_\mathbf{e}}{2}$ *we have:*

$$\mathbf{l}_{M(X,d_X)}(A_Y) \leq \mathbf{l}_{M(X,d_X)}(A_X) + K_\mathbf{c} n^2 (1 - e^{-\epsilon}) + K_\mathbf{e} n^2 (e^{\epsilon} - 1) \tag{2}$$

If $e_{M(X,d_X)(a)_{ij}}(x)$ *is constant in* x *(such as for any* M *that factors as* $M = MDS \circ H$*) we have:*

$$\mathbf{l}_{M(X,d_X)}(A_Y) \leq \mathbf{l}_{M(X,d_X)}(A_X) + K_\mathbf{c} n^2 (1 - e^{-\epsilon}) \tag{3}$$

(Proof in Appendix B.8)

These bounds apply to a very general class of manifold learning algorithms, including topologically unstable algorithms like IsoMap (Balasubramanian, 2002). As an example, consider using IsoMap to project n evenly spaced points that lie upon the surface of a radius r circle in $\mathbb{R}^2$ onto $\mathbb{R}^1$. In this case (X, d_X) is a finite ordered n-element subspace of $\mathbb{R}^2$ with Euclidean distance, $M = MDS \circ IsoCluster_\delta$ and for any matrix $A_X \in \mathbf{Mat}_{n,1}$ that consists of n evenly spaced points along the real line such that $A_{X_{i+1}} - A_{X_i} = 2r sin(\frac{2\pi}{2n})$ we have $\mathbf{l}_{M(X,d_X)}(A_X) = 0$. Now suppose that we instead apply IsoMap to a noisy view of (X, d_X): a finite ordered n-element subspace (Y, d_Y) of $\mathbb{R}^2$ where d_Y is Euclidean distance and $\forall_{i=1 \ldots n} d_X(X_i, Y_i) = d_Y(X_i, Y_i) = \|X_i - Y_i\| \leq \epsilon$. Then for any matrix $A_Y \in \mathbf{Mat}_{n,1}$ that minimizes $\mathbf{l}_{M(Y,d_Y)}$, Proposition 8 bounds the average squared difference between $|A_{Y_{i+1}} - A_{Y_i}|$ and $2r sin(\frac{2\pi}{2n})$.

3 Conclusion

We have taken the first steps towards a categorical framework for manifold learning. By defining a manifold learning algorithm as a functor out of a category of metric spaces, we can explicitly express the kind of dataset transformation that it is invariant to. We can also use functoriality to extend theorems about clustering algorithms to theorems about manifold learning algorithms, reason about our algorithms' stability properties, and create new algorithms that are guaranteed to obey these properties by composing functors. In future work we hope to use these techniques to derive more powerful theorems around the resilience of other kinds of unsupervised algorithms to noise.

References

Hervé Abdi. Metric multidimensional scaling (MDS): analyzing distance matrices. *Encyclopedia of Measurement and Statistics*, 2007.

Mukund Balasubramanian. The isomap algorithm and topological stability. *Science*, 295(5552): 7–7, 2002.

Andrew J Blumberg and Michael Lesnick. Universality of the homotopy interleaving distance. *arXiv preprint arXiv:1705.01690*, 2017.

Peter Bubenik and Jonathan A Scott. Categorification of persistent homology. *Discrete & Computational Geometry*, 51(3):600–627, 2014.

Gunnar Carlsson and Facundo Mémoli. Classifying clustering schemes. *Foundations of Computational Mathematics*, 2013.

Frédéric Chazal, Vin De Silva, and Steve Oudot. Persistence stability for geometric complexes. *Geometriae Dedicata*, 173(1):193–214, 2014.

Jared Culbertson, Dan P Guralnik, Jakob Hansen, and Peter F Stiller. Consistency constraints for overlapping data clustering. *arXiv preprint arXiv:1608.04331*, 2016.

Geoffrey Hinton. Non-linear dimensionality reduction. *University of Toronto Lecture Notes*, 2013. URL `https://www.cs.toronto.edu/~hinton/csc2535/notes/lec11new.pdf`.

Leland McInnes, John Healy, and James Melville. UMAP: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.

Luis N Scoccola. Locally persistent categories and metric properties of interleaving distances. *PhD Thesis at Western University*, 2020.

Dan Shiebler. Functorial clustering via simplicial complexes. *Topological Data Analysis and Beyond*, 2020.

Joshua B Tenenbaum, Vin De Silva, and John C Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.

A Appendix A: Experiment in Functorial Recombination

One benefit of the functorial perspective on manifold learning is that it provides a natural way to produce new manifold learning algorithms by recombining the components of existing algorithms. Suppose we are able to express two existing manifold learning algorithms M_1, M_2 in this framework such that $M_1 = L_1 \circ H_1$ and $M_2 = L_2 \circ H_2$ where H_1, H_2 are hierarchical clustering functors. Then we can use the compositionality of functors to define the manifold learning algorithms $L_2 \circ H_1$ or $L_1 \circ H_2$. We can use this procedure to combine the strengths of multiple algorithms in a way that preserves functoriality.

Consider a DNA recombination task in which we attempt to match a string of DNA that has been repeatedly mutated back to the original string. One way to solve this task is to generate embeddings for each string of DNA and look at the nearest neighbors of the mutated string. We can simulate this task as follows

1. Generate N original random sequences of DNA of length L (strings of "A", "C", "G", "T").
2. For each sequence, mutate the sequence M times to produce a mutation list, or a list of sequences which each start with an original DNA sequence and end with a final DNA sequence.
3. Collect each of the M sequences in each of these N mutation lists into a $N * M$ element finite pseudometric space with Hamming distance.
4. Build embeddings from this pseudometric space and compute the percent of mutation lists for which the nearest neighbor of the last DNA sequence in that list among the set of all original sequences is the first sequence in that list (the accuracy).

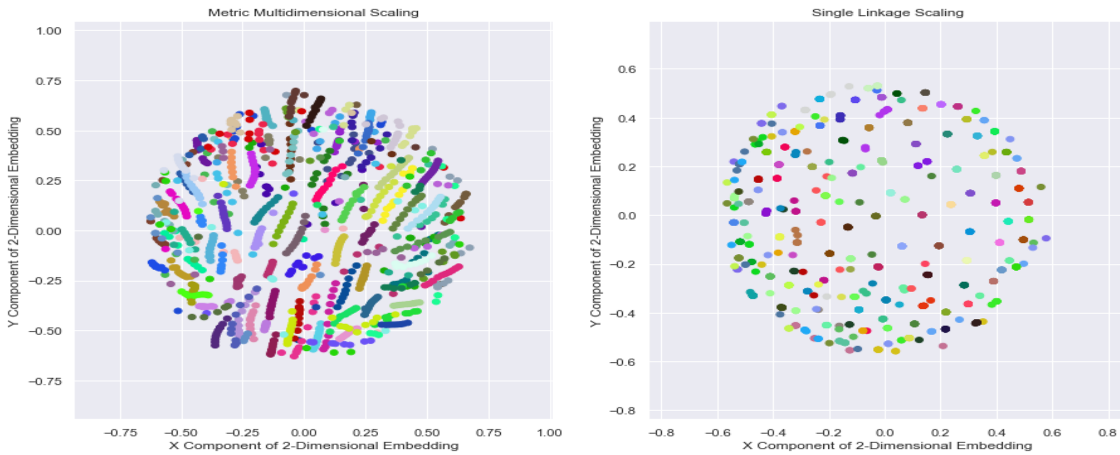

Figure 1: Embeddings of DNA sequences from the DNA recombination task with $L = 1000, N = 100, M = 10$. Each color indicates a unique DNA sequence mutation list. Note that Single Linkage Scaling ($MDS \circ \mathcal{SL}$) on the right embeds sequences in the same mutation list more closely together than Metric Multidimensional Scaling ($MDS \circ \mathcal{ML}$) on the left.

A manifold learning algorithm that performs well on this task will need to take advantage of the intermediate mutation states to recognize that the first state and final state in a mutation list should be embedded as close together as possible. We can follow the procedure in Section 1.1 to adapt the Metric Multidimensional Scaling algorithm $MDS \circ \mathcal{ML}$ (Section 1.2.1) into such an algorithm by forming the maximally interconnected functor $MDS \circ \mathcal{SL}$. Intuitively, this functor maps (X, d_X) to a loss function that corresponds to the optimization objective for Metric Multidimensional Scaling where Euclidean distance is replaced with:

$$d^*_X(x, x') = \inf\{\delta \mid \exists x = x_1, x_2, ..., x_n = x' \in X, d_X(x_i, x_{i+1}) \leq \delta\}$$

We call this the **Single Linkage Scaling** algorithm (Algorithm 1). Since this algorithm embeds points that are connected by a sequence in the original space as close together as possible, we expect Single Linkage Scaling to outperform Metric Multidimensional Scaling on the DNA recombination task. This is exactly what we see in Table 1. We also show the embeddings for each sequence in each list in Figure 1.

Algorithm 1 Single Linkage Scaling

```
procedure SingleLinkageScaling(((X, d_X), m))
    Initialize the |X| × |X| matrix B to all zeros
    ∀i, j ≤ |X|
        B_ij = inf{δ | ∃x_i = x_1, x_2, ..., x_n = x_j ∈ X, d_X(x_k, x_{k+1}) ≤ δ}
    A ← min_{A ∈ Mat_{|X|,m}} Σ_{i∈1...|X|, j∈1...|X|} (||A_i − A_j|| − B_ij)^2
    return A
```

B Appendix B: Proofs

B.1 Proof of Proposition 1

Proof. Recall the maximal linkage hierarchical overlapping clustering algorithm $\mathcal{ML}$ that maps the pseudometric space (X, d_X) to the fuzzy non-nested cover $(X, \mathcal{C}_{X_a})$ where $\mathcal{C}_{X_a}$ is the maximally linked sets of the relation R in which $x_1 R x_2$ if $d_X(x_1, x_2) \leq -log(a)$ (Shiebler, 2020; Culbertson et al., 2016). Consider also the function *Real* that maps the fuzzy non-nested cover $(X, \mathcal{C}_{X_a})$ to the pseudometric space (X, d'_X) in which:

$$d'_X(x_1, x_2) = e^{-\sup\{a \mid \exists S \in \mathcal{C}_{X_a}, x_1, x_2 \in S\}}$$

It is easy to see that $Real \circ \mathcal{ML}$ is an isometry on pseudometric spaces. Therefore, for any isometry-invariant manifold learning problem M, the composition $(M \circ Real) \circ \mathcal{ML}$ will have the same solution set as M. □

Algorithm	Accuracy $N = 100$ $M = 10$	Accuracy $N = 100$ $M = 20$	Accuracy $N = 200$ $M = 10$	Accuracy $N = 200$ $M = 20$
Metric Multidimensional Scaling Embedding Size 2	0.21 (± 0.05)	0.01 (± 0.02)	0.29 (± 0.02)	0.01 (± 0.00)
Single Linkage Scaling Embedding Size 2	0.61 (± 0.02)	0.68 (± 0.05)	0.76 (± 0.01)	0.32 (± 0.02)
Metric Multidimensional Scaling Embedding Size 5	0.74 (± 0.01)	0.13 (± 0.02)	0.84 (± 0.01)	0.04 (± 0.01)
Single Linkage Scaling Embedding Size 5	0.93 (± 0.05)	0.91 (± 0.02)	0.96 (± 0.02)	0.34 (± 0.02)

Table 1: Performance on the DNA recombination task of the Metric Multidimensional Scaling ($MDS \circ \mathcal{ML}$) and Single Linkage Scaling ($MDS \circ \mathcal{SL}$) algorithms. The accuracy is the percent of the N mutation lists of length M for which the nearest neighbor of the last sequence in the list among the set of all original DNA sequences is the first sequence in that list. The reported numbers are means (and standard deviations) across 10 simulations. All DNA sequences are of length $L = 1000$.

B.2 Proof of Proposition 2

Proof. By Proposition 6 in Shiebler (2020), there exist natural transformations from:

$$\mathcal{ML}(X, d_X)(W_H(-)) \to H(X, d_X)(-) \to \mathcal{SL}(X, d_X)(W_H(-))$$

where $W_H(a) = e^{-\delta_{H,a}}$. The statement then holds by functoriality. □

B.3 Proof of Proposition 3

Proof. Define $MDS : \mathbf{FCov}_{sur} \to \mathbf{FLoss}_m$ to map the fuzzy non-nested cover $H : (0, 1]^{op} \to \mathbf{Cov}_{inj}$ with vertex set X to $F : (0, 1]^{op} \to \mathbf{Loss}_m$ where $F(a) = (|X|, \{c_{F(a)_{ij}}, e_{F(a)_{ij}}\})$ and:

$$c_{F(a)_{ij}}(x) = x^2 + 2x^2 \begin{cases} 0 & \exists S \in H(a),\ x_i, x_j \in S \}) \\ 1/W_{ij} - 1/a & \text{else} \end{cases}$$

$$e_{F(a)_{ij}}(x) = \begin{cases} 0 & \exists S \in H(a),\ x_i, x_j \in S \}) \\ \frac{2log(W_{ij})}{W_{ij}} - \frac{2log(a)}{a} & \text{else} \end{cases}$$

where:

$$W_{ij} = \sup_{\geq 0}\{a \mid a \in (0, 1],\ \exists S \in H(a),\ x_i, x_j \in S\}$$

We will show that $MDS \circ \mathcal{ML}$ is an $\mathbf{PMet}_{sur}$-manifold learning functor that maps any pseudometric space (X, d_X) to the Metric Multidimensional Scaling embedding optimization problem over the distance matrix of d_X.

First, we need to show that $MDS : \mathbf{FCov}_{sur} \to \mathbf{FLoss}_m$ is a functor. Consider the fuzzy non-nested covers H_X and $H_{X'}$ in $\mathbf{FCov}_{sur}$ with vertex sets X, X' respectively such that there exists a morphism f in $\mathbf{FCov}_{sur}$ between them (a natural transformation with surjective components). Say $MDS(H_X)(a) = (|X|, \{c_{F(a)_{ij}}, e_{F(a)_{ij}}\})$ and $MDS(H_{X'})(a) = (|X'|, \{c'_{F(a)_{ij}}, e'_{F(a)_{ij}}\})$. Since each component of f is surjective it must be that $|X|' \leq |X|$. There are now two cases:

- Say $i, j \leq |X'|$. For each $a \in (0, 1], x_1, x_2 \in X, \exists S \in H_X(a),\ x_1, x_2 \in S$, by definition $\exists S \in H_{X'}(a),\ f(x_1), f(x_2) \in S$. Therefore $\forall_{x \in \mathbb{R}_{\geq 0}} e_{F(a)_{ij}}(x) \leq e'_{F(a)_{ij}}(x), c'_{F(a)_{ij}}(x) \leq c_{F(a)_{ij}}(x)$.
- Say $i > |X'|$ or $j > |X'|$. By definition $c'_{F(a)_{ij}}(x) = e'_{F(a)_{ij}}(x) = 0$. Since $c_{F(a)_{ij}}$ is non-negative and $e_{F(a)_{ij}}$ is non-positive, we have $\forall_{x \in \mathbb{R}_{\geq 0}} e_{F(a)_{ij}}(x) \leq e'_{F(a)_{ij}}(x), c'_{F(a)_{ij}}(x) \leq c_{F(a)_{ij}}(x)$.

Therefore $MDS(H_X) \leq MDS(H_{X'})$. Since $MDS : \mathbf{FCov}_{sur} \to \mathbf{FLoss}_m$ trivially preserves the identity we can conclude that it is a functor and $MDS \circ \mathcal{ML}$ is a $\mathbf{PMet}_{sur}$-manifold learning functor.

Next, we show that $MDS \circ \mathcal{ML}$ maps (X, d_X) to the Metric Multidimensional Scaling embedding optimization problem. Define $F = (MDS \circ \mathcal{ML})(X, d_X)$ and note that:

$$W_{ij} = \sup_{\geq 0}\{a \mid a \in (0,1],\ \exists S \in \mathcal{ML}(X, d_X)(a),\ x_i, x_j \in S\} = e^{-d_X(x_i,x_j)}$$

The F-loss is as follows:

$$\mathbf{l}_F(A) = \sum_{\substack{i \in 1...n \\ j \in 1...n}} \int_{a \in (0,1]} e_{F(a)_{ij}}(\|A_i - A_j\|) + c_{F(a)_{ij}}(\|A_i - A_j\|)\, da =$$

$$\sum_{\substack{i \in 1...n \\ j \in 1...n}} \|A_i - A_j\|^2 + \left(\int_{a \in (W_{ij},1]} \frac{2log(W_{ij})}{W_{ij}} - \frac{2log(a)}{a}\, da \right) +$$

$$\left(2\|A_i - A_j\|^2 \int_{a \in (W_{ij},1]} \frac{1}{W_{ij}} - \frac{1}{a}\, da \right) =$$

$$C + \sum_{\substack{i \in 1...n \\ j \in 1...n}} \|A_i - A_j\|^2 + log(W_{ij})^2 + 2\|A_i - A_j\|^2 log(W_{ij}) =$$

$$C + \sum_{\substack{i \in 1...n \\ j \in 1...n}} \|A_i - A_j\|^2 + d_X(x_i, x_j)^2 - 2\|A_i - A_j\|^2 d_X(x_i, x_j) =$$

$$C + \sum_{\substack{i \in 1...n \\ j \in 1...n}} \left(\|A_i - A_j\| - d_X(x_i, x_j) \right)^2$$

where $n = |X|$ and C is a constant factor.

□

B.4 Proof of Proposition 4

Proof. First, define the δ-graph of (X, d_X) to be the graph in which the vertices are the points in X and there exists an edge of length $d_X(x, x')$ between each pair of points $(x, x') \in X$ with $d_X(x, x') \leq \delta$. Now define $IsoCluster_\delta : \mathbf{PMet} \to \mathbf{FCov}$ such that $IsoCluster_\delta(X, d_X)(a)$ is the collection of maximally linked sets of the relation R_a, where for $x, x' \in X$ we have xR_ax' if there exists a path of length no larger than $-log(a)$ in the δ-graph of (X, d_X). We will show that $IsoCluster_\delta$ is a hierarchical **PMet**-clustering functor.

Consider the non-expansive map $f : (X, d_X) \to (Y, d_Y)$ and say that for some $a \in (0,1], x, x' \in X, \exists S \in IsoCluster_\delta(X, d_X)(a), x, x' \in S$. Then there exists $x = x_1, x_2, ..., x_{n-1}, x_n = x'$ such that:

$$\max_{i=1...n} d_X(x_i, x_{i+1}) \leq \delta \qquad \sum_{i=1...n} d_X(x_i, x_{i+1}) \leq -log(a)$$

which implies that:

$$\max_{i=1...n} d_Y(f(x_i), f(x_{i+1})) \leq \delta \qquad \sum_{i=1...n} d_Y(f(x_i), f(x_{i+1})) \leq -log(a)$$

which implies that $\exists S' \in IsoCluster(Y, d_Y)(a), f(x), f(x') \in S'$. Since $IsoCluster_\delta$ trivially preserves the identity and acts as the identity on the underlying set, we can conclude that $IsoCluster_\delta$ is a hierarchical **PMet**-clustering functor.

Next, we will show that the manifold learning functor $MDS \circ IsoCluster_\delta$ maps (X, d_X) to the IsoMap embedding optimization problem. First define:

$$W_{ij} = \sup_{\geq 0}\{a \mid a \in (0,1],\ \exists S \in IsoCluster_\delta(X, d_X)(a),\ x_i, x_j \in S\} = e^{-d'_X(x_i,x_j)}$$

where $d'_X(x_i, x_j)$ is the smallest γ such that there exists a path of length no greater than γ between x_i and x_j in the δ-graph of (X, d_X). Now if we define $F = (MDS \circ IsoCluster_\delta)(X, d_X)$ then following the same steps as in Section B.3 we have:

$$\mathbf{l}_F(A) = C + \sum_{\substack{i \in 1...n \\ j \in 1...n}} \left(\|A_i - A_j\| - d'_X(x_i, x_j) \right)^2$$

where $n = |X|$ and C is a constant factor. □

B.5 Proof of Proposition 5

Before we begin, we will show the following:

Proposition 9. *The category of fuzzy simplicial complexes and bijective simplicial maps* $\mathbf{FSCpx}_{bij}$ *(Shiebler, 2020) is finitely co-complete.*

Proof. For some finite category $\mathbf{C}$ consider a functor of the form $F : \mathbf{C} \to \mathbf{FSCpx}_{bij}$. Define the fuzzy simplicial complex $F_c : (0, 1]^{op} \to \mathbf{SCpx}_{bij}$ in $\mathbf{FSCpx}_{bij}$ to map $a \in (0, 1]$ to the simplicial complex whose set of n-simplices is $\cup_{o \in ob(\mathbf{C})} F(o)[n]$. Note that $ob(\mathbf{C})$ is the set of objects in $\mathbf{C}$. It is clear that this is the minimal fuzzy simplicial complex such that there exists a natural transformation from each fuzzy simplicial complex $F(o), o \in ob(\mathbf{C})$ into this fuzzy simplicial complex, so F_c is the colimit of F and $\mathbf{FSCpx}_{bij}$ is finitely co-complete. □

Now we will prove Proposition 5.

Proof. Note that for any $N, N' \in \mathbb{N}$ such that $N \neq N'$, the size N pseudometric spaces and the size N' pseudometric spaces have no morphisms between them in $\mathbf{PMet}_{isom}$. Therefore, we can uniquely define $FuzzySimplex$ by defining a separate functor $FuzzySimplex_N : \mathbf{PMet}_{isom}(N) \to \mathbf{FCov}_{bij}$ for each $N \in \mathbb{N}$, where $\mathbf{PMet}_{isom}(N)$ is the subcategory of $\mathbf{PMet}_{isom}$ where objects are restricted to pseudometric spaces (X, d_X) with cardinality N.

To start, denote the N-element discrete category $\mathbf{N}$ and define the following functor for step 1 (build a local uber-metric space around each point): $LocalMetric_N : \mathbf{PMet}_{isom}(N) \to \mathbf{UMet}^{\mathbf{N}}_{bij}$ sends the N-element pseudometric space (X, d_X) to the functor $F : \mathbf{N} \to \mathbf{UMet}_{bij}$ that maps $i \in \mathbf{N}$ to (X, d_{x_i}) where:

$$d_{x_i}(x_j, x_k) = \begin{cases} d_X(x_j, x_k) - min_{l=1...n} d_X(x_i, x_l) & i = j, i = k \\ \infty & else \end{cases}$$

$LocalMetric_N$ sends the function f to the natural transformation in which each component is f. Since f is an isometry this map must exist and be natural.

Since $LocalMetric_N$ trivially preserves composition and the identity it is a functor. For step 2 (convert each local uber-metric space to a fuzzy simplicial complex), we will use the functor (Shiebler, 2020):

$$(FinSing \circ -)_N : \mathbf{UMet}^{\mathbf{N}}_{bij} \to \mathbf{FSCpx}^{\mathbf{N}}_{bij}$$

which maps the functor $F : \mathbf{N} \to \mathbf{UMet}_{bij}$ to the functor $(FinSing \circ F) : \mathbf{N} \to \mathbf{FSCpx}_{bij}$.

For step 3 (take a fuzzy union of these fuzzy simplicial complexes), we apply the colimit functor $colim_N : \mathbf{FSCpx}^{\mathbf{N}}_{bij} \to \mathbf{FSCpx}_{bij}$ which sends an indexed set of fuzzy simplicial complexes in $\mathbf{FSCpx}^{\mathbf{N}}_{bij}$ to its logical fuzzy union. This functor exists by Proposition 9. In a logical fuzzy union the strength of a simplex is defined to be its maximum strength among the complexes we are adjoining[1] . For step 4 (convert the resulting fuzzy simplicial complex to a fuzzy non-nested flag cover), we use the functor $(Flag \circ -)$ from Shiebler (2020). Since $Flag$ maps bijective simplicial maps to bijections, the image of this functor over $\mathbf{FSCpx}_{bij}$ is $\mathbf{FCov}_{bij}$. Now we can compose steps 1-4 and apply a coproduct over $N \in \mathbb{N}$ to extend this to the following functor from $\mathbf{PMet}_{isom}$ to $\mathbf{FCov}_{bij}$:

$$FuzzySimplex = \Pi_{N \in \mathbb{N}} \, (Flag \circ -) \circ colim_N \circ (FinSing \circ -)_N \circ LocalMetric_N$$

We now show $FuzzySimplex$ is a hierarchical $\mathbf{PMet}_{isom}$-clustering functor. Since $FuzzySimplex$ is by definition a functor, we simply need to show for any (X, d_X) that $FuzzySimplex(X, d_X)$ is a fuzzy non-nested flag cover of X. First note that for any object $o \in \mathbf{N}$, the vertex set of the following fuzzy simplicial complex is X:

$$((FinSing \circ -)_N \circ LocalMetric_N)(X, d_X)(o)$$

Therefore the vertex set of the following fuzzy simplicial complex is X as well:

$$(colim_N \circ (FinSing \circ -)_N \circ LocalMetric_N)(X, d_X)$$

This implies that $FuzzySimplex(X, d_X)$ is a fuzzy cover of X.

□

[1]This is different from the probabilistic simplicial complex union that the UMAP python code uses (McInnes et al., 2018).

B.6 Proof of Proposition 6

Proof. Define $FCE : \mathbf{FCov}_{bij} \to \mathbf{FLoss}_m$ to map the fuzzy non-nested cover $H : (0,1]^{op} \to \mathbf{Cov}_{bij}$ with vertex set X to $F : (0,1]^{op} \to \mathbf{Loss}_m$ where $F(a) = (|X|, \{c_{F(a)_{ij}}, e_{F(a)_{ij}}\}, 0)$ and:

$$e_{F(a)_{ij}}(x) = \begin{cases} -log(e^{-x}) & \exists S \in H(a),\ x_i, x_j \in S\}) \\ 0 & \text{else} \end{cases}$$
$$c_{F(a)_{ij}}(x) = \begin{cases} 0 & \exists S \in H(a),\ x_i, x_j \in S\}) \\ -log(1 - e^{-x}) & \text{else} \end{cases}$$

We will show that $FCE \circ FuzzySimplex$ is an $\mathbf{PMet}_{isom}$-manifold learning functor that maps any pseudometric space (X, d_X) to the UMAP embedding optimization problem over the distance matrix of d_X.

First, we need to show that $FCE : \mathbf{FCov}_{bij} \to \mathbf{FLoss}_m$ is a functor. Consider the fuzzy non-nested covers H_X and $H_{X'}$ in $\mathbf{FCov}_{bij}$ with vertex sets X, X' respectively such that there exists a morphism f in $\mathbf{FCov}_{bij}$ between them (a natural transformation with bijective components). Say $FCE(H_X)(a) = (|X|, \{c_{F(a)_{ij}}, e_{F(a)_{ij}}\}, 0)$ and $FCE(H_{X'})(a) = (|X'|, \{c'_{F(a)_{ij}}, e'_{F(a)_{ij}}\}, 0)$. Since each component of f is bijective it must be that $|X|' = |X|$. Now for each $a \in (0,1], x_1, x_2 \in X, \exists S \in H_X(a),\ x_1, x_2 \in S$, by definition $\exists S \in H_{X'}(a),\ f(x_1), f(x_2) \in S$. Therefore $\forall_{x \in \mathbb{R}_{\geq 0}} e_{F(a)_{ij}}(x) \leq e'_{F(a)_{ij}}(x), c'_{F(a)_{ij}}(x) \leq c_{F(a)_{ij}}(x)$. Therefore $FCE(H_X) \leq FCE(H_{X'})$. Since $FCE : \mathbf{FCov}_{bij} \to \mathbf{FLoss}_m$ trivially preserves the identity we can conclude that it is a functor and $FCE \circ FuzzySimplex$ is a $\mathbf{PMet}_{isom}$-manifold learning functor.

Next, we will show that $FCE \circ FuzzySimplex$ maps (X, d_X) to the UMAP embedding optimization problem. Define $F = (FCE \circ FuzzySimplex)(X, d_X)$. We have that the F-loss is:

$$\mathbf{l}_F(A) = \sum_{\substack{i \in 1...n \\ j \in 1...n}} \int_{a \in (0,1]} c_{F(a)_{ij}}(\|A_i - A_j\|) + e_{F(a)_{ij}}(\|A_i - A_j\|) da =$$

$$\sum_{\substack{i \in 1...n \\ j \in 1...n}} \int_{a \in (W_{ij}, 1]} -log(1 - e^{-\|A_i - A_j\|})\, da - \int_{a \in (0, W_{ij}]} log(e^{-\|A_i - A_j\|})\, da =$$

$$\sum_{\substack{i \in 1...n \\ j \in 1...n}} -(1 - W_{ij}) log(1 - e^{-\|A_i - A_j\|}) - W_{ij} log(e^{-\|A_i - A_j\|}) =$$

$$C + \sum_{\substack{i \in 1...n \\ j \in 1...n}} (1 - W_{ij}) log\left(\frac{1 - W_{ij}}{1 - e^{-\|A_i - A_j\|}}\right) + W_{ij} log\left(\frac{W_{ij}}{e^{-\|A_i - A_j\|}}\right)$$

where $W_{ij} = \sup_{\geq 0}\{a \mid a \in (0,1],\ \exists S \in FuzzySimplex(X, d_X)(a),\ x_i, x_j \in S\}$ is the weight of the fuzzy 1-simplex connecting x_i and x_j, $n = |X|$ and $C = \sum_{\substack{i \in 1...n \\ j \in 1...n}} (1 - W_{ij}) log(1 - W_{ij}) + W_{ij} log(W_{ij})$ is a constant. □

B.7 Proof of Proposition 7

Proof. Say we have a pair of finite metric spaces $(X, d_X), (Y, d_Y)$ such that there exists a pair of morphisms $f : (X, d_X) \to (Y, d_Y + \epsilon), g : (Y, d_Y) \to (X, d_X + \epsilon)$ in $\mathbf{D}$. By definition we have that $H(X, d_X + \epsilon)(-log(\delta)) = H(X, d_X)(-log(\delta + \epsilon))$, so by functoriality for any $\delta \in \mathbb{R}_{\geq 0}$ we have that f is refinement-preserving from $H(X, d_X)(-log(\delta))$ to $H(Y, d_Y)(-log(\delta + \epsilon))$ and g is refinement-preserving from $H(Y, d_Y)(-log(\delta))$ to $H(X, d_X)(-log(\delta + \epsilon))$. Therefore since $M = L \circ H$ by functoriality we also have that:

$$M(X, d_X)(-log(\delta)) \leq M(Y, d_Y)(-log(\delta + \epsilon))$$
$$M(Y, d_Y)(-log(\delta)) \leq M(X, d_X)(-log(\delta + \epsilon))$$

Since $\mathbf{Loss}_m$ is a preorder, this implies that $M(X, d_X)$ and $M(Y, d_Y)$ are ϵ-interleaved. □

B.8 Proof of Proposition 8

To start, we will show the following:

Proposition 10. *Suppose M is a standard form $\mathbf{PMet}_{sur}$-manifold learning functor and M' is a standard form $\mathbf{PMet}_{inj}$-manifold learning functor. Then for any (X, d_X) and $a \in (0, 1]$ we have that $e_{M(X,d_X)(a)_{ij}}, c_{M'(X,d_X)(a)_{ij}}$ are non-positive and $c_{M(X,d_X)(a)_{ij}}, e_{M'(X,d_X)(a)_{ij}}$ are non-negative.*

Proof. First, since there trivially exists a surjective non-expansive map from (X, d_X) to $(\{*\}, 0)$, by functoriality we have that $M(X, d_X) \leq M(\{*\}, 0)$. This implies that for all i, j we have $e_{M(X,d_X)(a)_{ij}} \leq e_{M(\{*\},0)(a)_{ij}} = 0$ and $0 = c_{M(\{*\},0)(a)_{ij}} \leq c_{M(X,d_X)(a)_{ij}}$.

Next, since there trivially exists an injective non-expansive map from $(\{*\}, 0)$ to (X, d_X), by functoriality we have that $M'(\{*\}, 0) \leq M'(X, d_X)$. This implies that for all i, j we have $c_{M'(X,d_X)(a)_{ij}} \leq c_{M'(\{*\},0)(a)_{ij}} = 0$ and $0 = e_{M'(\{*\},0)(a)_{ij}} \leq e_{M'(X,d_X)(a)_{ij}}$. □

We can now proceed with the proof of Proposition 8:

Proof. By using Proposition 7, we see that in order to prove Proposition 8 we simply need to show that if F, G are ϵ^*-interleaved functors in $\mathbf{FLoss}_m$ such that $A_F \in \mathbf{Mat}_{n,m}$ minimizes $\mathbf{l}_F$, $A_G \in \mathbf{Mat}_{n,m}$ minimizes $\mathbf{l}_G$, $c_{F(a)_{ij}}, c_{G(a)_{ij}}$ are non-negative, $e_{F(a)_{ij}}, e_{G(a)_{ij}}$ are non-positive, $|c_{F(a)_{ij}}(x)| \leq \frac{K_\mathbf{c}}{2}$, $|c_{G(a)_{ij}}(x)| \leq \frac{K_\mathbf{c}}{2}$ and $|e_{F(a)_{ij}}(x)| \leq \frac{K_\mathbf{e}}{2}$, $|e_{G(a)_{ij}}(x)| \leq \frac{K_\mathbf{e}}{2}$ then we have:

$$\mathbf{l}_F(A_G) \leq \mathbf{l}_F(A_F) + K_\mathbf{c} n^2(1 - e^{-\epsilon}) + K_\mathbf{e} n^2(e^{\epsilon} - 1)$$

And that in the special case where $c_{F(a)_{ij}}(x)$ is constant in x we have:

$$\mathbf{l}_F(A_G) \leq \mathbf{l}_F(A_F) + K_\mathbf{c} n^2(1 - e^{-\epsilon})$$

Now for simplicity we will write:

$$\mathbf{e}_{F(a)}(A) = \sum_{\substack{i \in 1...n \\ j \in 1...n}} e_{F(a)_{ij}}(\|A_i - A_j\|) \qquad \mathbf{c}_{F(a)}(A) = \sum_{\substack{i \in 1...n \\ j \in 1...n}} c_{F(a)_{ij}}(\|A_i - A_j\|)$$

By the definition of ϵ-interleaving we have the following for any $A \in \mathbf{Mat}_{n,m}$.

$$\mathbf{c}_{F(d * e^{-\epsilon})}(x) \leq \mathbf{c}_{G(d)}(x) \qquad \mathbf{e}_{G(d)}(x) \leq \mathbf{e}_{F(d * e^{-\epsilon})}(x)$$

Now we can conclude that:

$$\begin{aligned}
\mathbf{l}_F(A_G) = \int_{a \in (0,1]} \mathbf{c}_{F(a)}(A_G)\, da + \int_{a \in (0,1]} \mathbf{e}_{F(a)}(A_G)\, da \leq^{**} \\
e^{-\epsilon} \int_{a \in (0,1]} \mathbf{c}_{G(a)}(A_G)\, da + \frac{K_\mathbf{c}}{2} n^2(1 - e^{-\epsilon}) + \int_{a \in (0,1]} \mathbf{e}_{F(a)}(A_G)\, da \leq \\
\int_{a \in (0,1]} \mathbf{c}_{G(a)}(A_G)\, da + \frac{K_\mathbf{c}}{2} n^2(1 - e^{-\epsilon}) + \int_{a \in (0,1]} \mathbf{e}_{F(a)}(A_G)\, da \leq^{***} \\
\int_{a \in (0,1]} \mathbf{c}_{G(a)}(A_G)\, da + e^{\epsilon} \int_{a \in (0,1]} \mathbf{e}_{G(a)}(A_G)\, da + \frac{K_\mathbf{c}}{2} n^2(1 - e^{-\epsilon}) + (e^{\epsilon} - 1)\frac{K_\mathbf{e}}{2} n^2 \leq \\
\left(\int_{a \in (0,1]} \mathbf{c}_{G(a)}(A_G) + \mathbf{e}_{G(a)}(A_G)\, da\right) + \frac{K_\mathbf{c}}{2} n^2(1 - e^{-\epsilon}) + (e^{\epsilon} - 1)\frac{K_\mathbf{e}}{2} n^2 \leq^{*} \\
\left(\int_{a \in (0,1]} \mathbf{e}_{G(a)}(A_F) + \mathbf{c}_{G(a)}(A_F)\, da\right) + \frac{K_\mathbf{c}}{2} n^2(1 - e^{-\epsilon}) + (e^{\epsilon} - 1)\frac{K_\mathbf{e}}{2} n^2 \leq^{**} \\
\int_{a \in (0,1]} \mathbf{e}_{G(a)}(A_F) + e^{-\epsilon}\mathbf{c}_{F(a)}(A_F)\, da + K_\mathbf{c} n^2(1 - e^{-\epsilon}) + (e^{\epsilon} - 1)\frac{K_\mathbf{e}}{2} n^2 \leq \\
\int_{a \in (0,1]} e^{\epsilon}\mathbf{e}_{F(a)}(A) + e^{-\epsilon}\mathbf{c}_{F(a)}(A_F)\, da + K_\mathbf{c} n^2(1 - e^{-\epsilon}) + (e^{\epsilon} - 1)K_\mathbf{e} n^2 \leq \\
\mathbf{l}_F(A_F) + K_\mathbf{c} n^2(1 - e^{-\epsilon}) + (e^{\epsilon} - 1)K_\mathbf{e} n^2
\end{aligned}$$

In the special case where $e_{F(a)_{ij}}$ is constant we have:

$$
\begin{aligned}
\mathbf{l}_F(A_G) = \int_{a\in(0,1]} \mathbf{c}_{F(a)}(A_G)\,da + \int_{a\in(0,1]} \mathbf{e}_{F(a)}(A_G)\,da \leq^{**} \\
e^{-\epsilon}\int_{a\in(0,1]} \mathbf{c}_{G(a)}(A_G)\,da + \frac{K_\mathbf{c}}{2}n^2(1-e^{-\epsilon}) + \int_{a\in(0,1]} \mathbf{e}_{F(a)}(A_G)\,da \leq^{*} \\
e^{-\epsilon}\int_{a\in(0,1]} \mathbf{c}_{G(a)}(A_F)\,da + \frac{K_\mathbf{c}}{2}n^2(1-e^{-\epsilon}) + \int_{a\in(0,1]} \mathbf{e}_{F(a)}(A_F)\,da \leq^{**} \\
e^{-2\epsilon}\int_{a\in(0,1]} \mathbf{c}_{F(a)}(A_F)\,da + K_\mathbf{c}n^2(1-e^{-\epsilon}) + \int_{a\in(0,1]} \mathbf{e}_{F(a)}(A_F)\,da \leq \\
\mathbf{l}_F(A_F) + K_\mathbf{c}n^2(1-e^{-\epsilon})
\end{aligned}
$$

The steps marked with $*$ hold by the optimality of A_G. The steps marked with $**$ are by the following, which holds because $\mathbf{c}$ is non-negative and increasing in a:

$$
\begin{aligned}
\int_{a\in(0,1]} \mathbf{c}_{F(a)}(A)\,da \leq \\
\int_{a\in(0,1]} \mathbf{c}_{F(a)}(A)\,da - \int_{a\in(e^{-\epsilon},1]} \mathbf{c}_{F(a)}(A)\,da + \frac{K_\mathbf{c}}{2}n^2(1-e^{-\epsilon}) = \\
\int_{a\in(0,e^{-\epsilon}]} \mathbf{c}_{F(a)}(A)\,da + \frac{K_\mathbf{c}}{2}n^2(1-e^{-\epsilon}) = \\
e^{-\epsilon}\int_{a\in(0,1]} \mathbf{c}_{F(a*e^{-\epsilon})}(A)\,da + \frac{K_\mathbf{c}}{2}n^2(1-e^{-\epsilon}) \leq \\
e^{-\epsilon}\int_{a\in(0,1]} \mathbf{c}_{G(a)}(A)\,da + \frac{K_\mathbf{c}}{2}n^2(1-e^{-\epsilon})
\end{aligned}
$$

The steps marked with $***$ are by the following, which holds because $\mathbf{e}$ is non-positive and decreasing in a:

$$
\begin{aligned}
\int_{a\in(0,1]} \mathbf{e}_{F(a)}(A)\,da \leq \\
\int_{a\in(0,1]} \mathbf{e}_{G(a*e^{-\epsilon})}(A)\,da = \\
\frac{1}{e^{-\epsilon}}\int_{a\in(0,e^{-\epsilon}]} \mathbf{e}_{G(a)}(A)\,da = \\
\frac{1}{e^{-\epsilon}}\left(\int_{a\in(0,1]} \mathbf{e}_{G(a)}(A)\,da + \int_{a\in(e^{-\epsilon},1]} \mathbf{e}_{G(a)}(A)\,da\right) \leq \\
\frac{1}{e^{-\epsilon}}\left(\int_{a\in(0,1]} \mathbf{e}_{G(a)}(A)\,da - (1-e^{-\epsilon})K_\mathbf{e}n^2\,da\right) = \\
e^{\epsilon}\int_{a\in(0,1]} \mathbf{e}_{G(a)}(A)\,da - (e^{\epsilon}-1)K_\mathbf{e}n^2
\end{aligned}
$$

$\square$