Functorial Manifold Learning

Dan Shiebler

We adapt previous research on category theory and topological unsupervised learning to develop a functorial perspective on manifold learning, also known as nonlinear dimensionality reduction. We first characterize manifold learning algorithms as functors that map pseudometric spaces to optimization objectives and that 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 equivariants. We express several popular manifold learning algorithms as functors at different levels of this hierarchy, including Metric Multidimensional Scaling, IsoMap, and UMAP. Next, we use interleaving distance to study the stability of a broad class of manifold learning algorithms. We present bounds on how closely the embeddings these algorithms produce from noisy data approximate the embeddings they would learn from noiseless data. Finally, we use our framework to derive a set of novel manifold learning algorithms, which we experimentally demonstrate are competitive with the state of the art.

1 Introduction

Suppose we have a finite pseudometric space (X, d_X) that we assume has been sampled from some larger space $\mathbf{X}$ according to some probability measure $\mu_{\mathbf{X}}$ over $\mathbf{X}$. **Manifold Learning** algorithms like Isomap [22], Metric Multidimensional Scaling [1], and UMAP [18] construct $\mathbb{R}^m$-embeddings for the points in X, which we interpret as coordinates for the support of $\mu_{\mathbf{X}}$. These techniques are based on the assumption that this support can be well-approximated with a manifold. In this paper we use **functoriality**, a basic concept from Category Theory, to explore two aspects of manifold learning algorithms:

- **Equivariance**: A manifold learning algorithm is equivariant to a transformation if applying that transformation to its inputs results in an corresponding transformation of its outputs. Understanding the equivariance of a transform lets us understand how it will behave on new types of data.
- **Stability**: The stability of a manifold learning algorithm captures how well the embeddings it generates on noisy data approximate the embeddings it would generate on noiseless data. Understanding the stability of a transform helps us predict its performance on real-world applications.

1.1 Functoriality

In order for a manifold learning algorithm to be useful, the properties of the embeddings that the algorithm derives from (X, d_X) must be somewhat in line with the structure of (X, d_X). One way to formalize this is to cast the algorithm as a functor between categories. A **category** is a collection of objects and morphisms between them. Morphisms are closed under an associative composition operation, and each object is equipped with an identity morphism. An example category is the collection of sets (objects) and functions (morphisms) between them.

A **functor** is a mapping between categories that preserves identity morphisms and morphism composition. Underlying this straightforward definition is a powerful concept: the functoriality of a transformation is a blueprint for its structure, expressed in terms of the equivariants it preserves. If a given transformation is functorial over some pair of categories, then the transformation preserves the structure represented in those categories' morphisms. By identifying the settings under which an algorithm is functorial, we can derive extensions of the algorithm that preserve functoriality and identify modifications that break it. See “Basic Category Theory” [17] or “Seven Sketches in Compositionality” [13] for more details on categories and functoriality.

1.2 Summary of Contributions

In Section 2, we demonstrate that manifold learning algorithms can be expressed as optimization problems defined on top of hierarchical overlapping clusterings. That is, we can express these algorithms in terms of the composition of:

- a strategy for clustering points at different distance scales
- an order-preserving transformation from a clustering of points to a loss function

We formalize this relationship in terms of the composition of functors between categories of pseudometric spaces, hierarchical overlapping clusterings, and optimization problems. This allows us to formally extend clustering theory into manifold learning theory.

In Section 2.1 we build on clustering theory to demonstrate that every manifold learning objective lies on a spectrum based on the criterion by which embeddings are penalized for being too close together or too far apart. In Section 2.2 we introduce a hierarchy of manifold learning algorithms and categorize algorithms based on the dataset transformations over which they are equivariant. In Section 2.3 we provide several examples of this categorization. We show that UMAP is equivariant to isometries and both IsoMap and Metric Multidimensional Scaling are equivariant to surjective non-expansive maps. Identifying these equivariants is helpful for identifying the circumstances under which we would expect our algorithms to generalize. For example, while adding points to a dataset will modify the IsoMap objective in a predictable way, this will completely disrupt the UMAP objective. This is caused by the fact that UMAP uses a local distance rescaling procedure that is density-sensitive, and is therefore not equivariant to injective or surjective non-expansive maps.

In Section 3 we use interleaving distance to study the stability of a broad class of manifold learning algorithms. We present novel bounds on how well the embeddings that algorithms in this class learn on noisy data approximate the embeddings they learn on noiseless data.

In Section 4 we build on this theory to describe a strategy for deriving novel algorithms from existing manifold learning algorithms. As an example we derive the **Single Linkage Scaling** algorithm, which projects samples in the same connected component of the Rips complex as close together as possible. We also present experimental results demonstrating the efficacy of this algorithm.

1.3 Related Work

Several authors have explored functorial perspectives on clustering algorithms. Carlsson et al. [8] introduce clustering functors that map metric spaces to partitionings, whereas Culbertson et al. [11] take a slightly broader scope and also explore overlapping clustering functors that map metric spaces to coverings. Both approaches demonstrate that metric space categories with fewer morphisms permit a richer class of clustering functors. For example, while the single linkage clustering algorithm is functorial over the full category of metric spaces and non-expansive maps, density-sensitive clustering algorithms like robust single linkage are only functorial over the subcategory of metric spaces and injective non-expansive maps. In order to get around the Kleinberg Impossibility Theorem [16], which states that any scale invariant flat clustering must sacrifice either surjectivity or a notion of consistency, several authors [8, 12, 19] also explore hierarchical clustering functors that map metric spaces to multi-scale dataset partitionings or covers. Shiebler [20] builds on this perspective to factor clustering functors through a category of simplicial complexes.

Manifold learning shares structure with hierarchical clustering, and some authors have begun applying categorical ideas to manifold learning. For example, McInnes et al. [18] introduce the UMAP manifold learning algorithm in terms of Spivak's fuzzy simplicial sets [21], which are a categorical analog of simplicial filtrations.

In Section 3 we study the stability of manifold learning algorithms to dataset noise. Due to the importance of this topic, many other authors have researched the stability properties of manifold learning algorithms. For example, Baily [3] explore adaptations of PCA to noisy data, and Gerber et al. [15] demonstrate that Laplacian Eigenmaps has nicer stability properties than IsoMap. However, we believe that ours is the first work that uses interleaving distance to formalize a stability property.

1.4 Preliminaries on Functorial Hierarchical Overlapping Clustering

We briefly review some definitions from the functorial perspective on hierarchical overlapping clustering. For more details, see Shiebler [20]. Given a set X, a **non-nested flag cover** $\mathcal{C}_X$ of X is a cover of X such that: (1) if $A,B \in \mathcal{C}_X$ and $A \subseteq B$, then $A = B$, (2) the simplicial complex with vertices corresponding to the elements of X and faces all finite subsets of the sets in $\mathcal{C}_X$ is a flag complex, or a simplicial complex that can be expressed as the clique of its 1-skeleton. The category **Cov** has tuples $(X, \mathcal{C}_X)$ as objects where $\mathcal{C}_X$ is a non-nested flag cover of the finite set X. The morphisms between $(X, \mathcal{C}_X)$ and $(Y, \mathcal{C}_Y)$ are functions $f : X \to Y$ where for any set S in $\mathcal{C}_X$ there exists some set S' in $\mathcal{C}_Y$ such that $f(S) \subseteq S'$.

Next, we will represent datasets with the category **PMet** of finite pseudometric spaces and non-expansive maps between them. Given a subcategory **D** of **PMet**, a **flat D-clustering functor** is a functor $C : \mathbf{D} \to \mathbf{Cov}$ that is the identity on the underlying set. Intuitively, a flat **D**-clustering functor maps a dataset (X, d_X) to a cover of the set X in a way such that increasing the distances between points in X may cause clusters to separate.

A **fuzzy non-nested flag cover** is a functor $F_X : (0,1]^{op} \to \mathbf{Cov}$ such that for any morphism $a \le a'$ in $(0,1]$, the morphism $F_X(a \le a')$ is the identity on the underlying set. In the category **FCov** objects are fuzzy non-nested covers and morphisms are natural transformations between them. Given a subcategory **D** of **PMet** a **hierarchical D-clustering functor** is a functor $H : \mathbf{D} \to \mathbf{FCov}$ such that for $a \in (0,1]$, $H(-)(a) : \mathbf{PMet} \to \mathbf{Cov}$ is a flat **D**-clustering functor. Intuitively, a hierarchical **D**-clustering functor maps a pair of a dataset (X, d_X) and a strength $a \in (0,1]$ to a cover of the set X in a way such that increasing the distances between points in X or increasing the strength a may cause clusters to separate.

2 Manifold Learning

A manifold learning algorithm constructs an $n \times m$ real valued matrix of embeddings in $\mathbf{Mat}_{n,m} = \mathbb{R}^{n \times m}$ 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. In particular, we focus on **pairwise embedding optimization problems**, or embedding optimization problems where l can be expressed as a sum of pairwise terms $l_{ij} : \mathbb{R}_{\ge 0} \to \mathbb{R}$ such that $l(A) = \sum_{\substack{i \in 1\ldots n \\ j \in 1\ldots n}} l_{ij}(\|A_i - A_j\|)$. We express such a pairwise embedding optimization problem with the tuple $(n, m, \{l_{ij}\})$.

Formally we define a **manifold learning problem** to be a function that maps the pseudometric space (X, d_X) to a pairwise embedding optimization problem of the form $(|X|, m, \{l_{ij}\})$. Note that this definition does not include any specification of how the optimization problem is solved (gradient descent, reduction to an eigenproblem, etc). For example, the Metric Multidimensional Scaling manifold learning problem maps the pseudometric space (X, d_X) to $(|X|, m, \{l_{ij}\})$ where $l_{ij}(\delta) = (d_X(x_i, x_j) - \delta)^2$. Optimizing this objective involves finding a matrix A that minimizes $\sum_{i,j \in 1...|X|}(d_X(x_i, x_j) - \|A_i - A_j\|)^2$. That is, the Euclidean distance matrix of the rows of the optimal 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. A particularly useful property of isometry-invariant manifold learning problems is that they factor through hierarchical clustering algorithms.

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 [20]) 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 A.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 pairwise embedding optimization problems and replace functions with functors from **PMet** into this category. To start, consider the category **L**:

Definition 2.1. *The objects in* **L** *are tuples $(n, \{l_{ij}\})$ where n is a natural number and $l_{ij} : \mathbb{R}_{\geq 0} \to \mathbb{R}$ is a real-valued function that satisfies $l_{i'j'}(x) = 0$ for $i' > n$ or $j' > n$.* **L** *is a preorder where $(n, \{l_{ij}\}) \leq (n', \{l'_{ij}\})$ iff for any $x \in \mathbb{R}_{\geq 0}, i, j \in \mathbb{N}$ we have $l_{ij}(x) \leq l'_{ij}(x)$.*

Given a choice of m, we can view the objects in **L** as pairwise embedding optimization problems. However, **L** is not quite expressive enough to serve as our target category. Recall the Metric Multidimensional Scaling manifold learning problem, which maps the pseudometric space (X, d_X) to the pairwise embedding optimization problem $(|X|, m, \{l_{ij}\})$ where $l_{ij}(\delta) = (d_X(x_i, x_j) - \delta)^2$. Since l_{ij} does not vary monotonically with d_X, it is clear that this manifold learning problem is not a functor from **PMet** to **L**. However, note that we can express $l_{ij}(A)$ as the sum of a term that increases monotonically in $d_X(x_i, x_j)$ and a term that decreases monotonically in $d_X(x_i, x_j)$:

$$l_{ij}(\delta) = (d_X(x_i, x_j) - \delta)^2 = \left(\delta^2 + d_X(x_i, x_j)^2\right) - \left(2\delta d_X(x_i, x_j)\right)$$

We will see in Section 2.3 that the embedding optimization problems associated with many common manifold learning algorithms decompose similarly. We can build on this insight to create a new category **Loss** with the following pullback:

$$\begin{tikzcd}
\mathbf{Loss} \arrow[r, dotted] \arrow[d, dotted] & \mathbf{L} \arrow[d, "U"] \\
\mathbf{L}^{op} \arrow[r, "U"] & \mathbb{N}
\end{tikzcd}$$

where U is the forgetful functor that maps $(n, \{l_{ij}\})$ to n. Intuitively **Loss** is a subcategory of $\mathbf{L}^{op} \times \mathbf{L}$ and we can write the objects in **Loss** as tuples $(n, \{c_{ij}, e_{ij}\})$ 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 $e_{ij}(x) \leq e'_{ij}(x)$. Given a choice of m, each object $(n, \{c_{ij}, e_{ij}\})$ in **Loss** corresponds to the pairwise embedding optimization problem $(n, m, \{l_{ij}\})$ where $l_{ij}(\delta) = c_{ij}(\delta) + e_{ij}(\delta)$.

Similarly to the representation of hierarchical clustering algorithms as maps into a category **FCov** of functors $(0, 1]^{\mathrm{op}} \to \mathbf{Cov}$, we will represent manifold learning algorithms as mapping into a category

FLoss of functors $(0,1]^{\text{op}} \to \mathbf{Loss}$. The objects in **FLoss** are functors $F : (0,1]^{\text{op}} \to \mathbf{Loss}$ that commute with the forgetful functor that maps $(n, \{c_{ij}, e_{ij}\})$ to n, and the morphisms are natural transformations. We call n the **cardinality** of F. We can define a functor $Flatten : \mathbf{FLoss} \to \mathbf{Loss}$ that maps the functor F where $F(a) = (n, \{c_{F(a)_{ij}}, e_{F(a)_{ij}}\})$ to the tuple $(n, \{c_{ij}, e_{ij}\})$ where:

$$c_{ij}(x) = \int_{a \in (0,1]} c_{F(a)_{ij}}(x)\, da \qquad e_{ij}(x) = \int_{a \in (0,1]} e_{F(a)_{ij}}(x)\, da$$

Therefore, each functor $F \in \mathbf{FLoss}$ with cardinality n corresponds to the pairwise embedding optimization problem $(n, m, \{l_{F_{ij}}\})$ where $l_{F_{ij}}(\delta) = \int_{a \in (0,1]} c_{F(a)_{ij}}(\delta) + e_{F(a)_{ij}}(\delta)\, da$. We will call the sum of these terms, $\mathbf{l}_F(A)$, the F-loss:

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

We can now give our definition of a manifold learning functor:

Definition 2.2. *Suppose* **PMet** *is the category of pseudometric spaces and non-expansive maps and* **FCov** *is the category of fuzzy non-nested flag covers and natural transformations (see Section 1.4). Then given the subcategories* $\mathbf{D} \subseteq \mathbf{PMet}, \mathbf{D}' \subseteq \mathbf{FCov}$*, the composition* $L \circ H : \mathbf{D} \to \mathbf{FLoss}$ *forms a* **D*-manifold learning functor*** *if* $H : \mathbf{D} \to \mathbf{D}'$ *is a hierarchical* **D***-clustering functor and* $L : \mathbf{D}' \to \mathbf{Loss}$ *is a functor that maps a fuzzy non-nested flag cover with vertex set* X *to some* $F_X \in \mathbf{FLoss}$ *with cardinality* $|X|$*.*

Intuitively a manifold learning functor $\mathbf{D} \xrightarrow{H} \mathbf{D}' \xrightarrow{L} \mathbf{FLoss}$ 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 in **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 $(|X|, m, \mathbf{l}_{M(X, d_X)})$.

2.1 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 flag 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 $d_X(x_1, x_2) \leq -\log(a)$ [20, 11]. If we apply functoriality to Proposition 6 in Shiebler [20] 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 [20, 11] 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 A.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 $L : \mathbf{PMet}_{inj} \to \mathbf{Loss}$ 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 family $\mathcal{L}_k$ of hierarchical overlapping clustering functors from Culbertson et al. [11]: 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_ax' 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 Section 4): 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 $L \circ \mathcal{L}_k$.

2.2 Characterizing Manifold Learning Problems

Similarly to how Carlsson et al. [8] 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 problem 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**. Intuitively, many positive-extensible manifold learning problems are minmax problems that aim to simultaneously minimize $|c|$ and maximize $|e|$. Any $\mathbf{PMet}_{sur}$-manifold learning functor is positive extensible and any $\mathbf{PMet}_{inj}$-manifold learning functor is negative extensible.

Proposition 3. *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 in Appendix A.3)*

In the next section we show examples of manifold learning algorithms in each of these categories.

2.3 Examples

2.3.1 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 [1] 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 [2]. Now define $MDS : \mathbf{FCov}_{sur} \to \mathbf{FLoss}$ to map the fuzzy non-nested flag cover $H : (0, 1]^{\text{op}} \to \mathbf{Cov}_{inj}$ with vertex set X to $F : (0, 1]^{\text{op}} \to \mathbf{Loss}$ where

$F(a) = (|X|, \{c_{F(a)_{ij}}, e_{F(a)_{ij}}\}, 0)$ and:

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

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

where:

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

Proposition 4. *MDS is a functor, and $MDS \circ \mathcal{ML}$ is a* $\mathbf{PMet}_{sur}$*-manifold learning functor that maps the finite pseudometric space (X, d_X) to the Metric Multidimensional Scaling embedding optimization problem. (Proof in Appendix A.4)*

2.3.2 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 [22] 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 5. *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 A.5)*

2.3.3 k-Path Scaling ($\mathbf{PMet}_{sur}$-Manifold Learning Functor)

Given a finite pseudometric space (X, d_X) and $k \in \mathbb{N}$, suppose $d'_X(x_i, x_j)$ is the smallest δ such that there exists a path of $\leq k$ edges between x_i and x_j in the δ-Vietoris Rips complex of (X, d_X). Then the $\mathbf{PMet}_{sur}$-manifold learning functor $MDS \circ \mathcal{L}_k$ maps (X, d_X) to the k-**Path Scaling** embedding optimization problem $(|X|, m, l)$ where $l(A) = \sum_{i,j \in 1...|X|} (d'_X(x_i, x_j) - \|A_i - A_j\|)^2$.

2.3.4 k-Vertex-Connected Scaling ($\mathbf{PMet}_{bij}$-Manifold Learning Functor)

For $k \in \mathbb{N}$ the hierarchical overlapping clustering functor $\mathcal{VL}_k$ maps the finite pseudometric space (X, d_X) to the fuzzy non-nested flag cover (X, C_{X_a}) where C_{X_a} is the set of maximal $\min(|X|, k)$-vertex-connected subgraphs of the $-\log(a)$-Vietoris-Rips complex of (X, d_X). Note that $\mathcal{VL}_1 = \mathcal{SL}$ and $lim_{k \to \infty} \mathcal{VL}_k = \mathcal{ML}$. Note also that for $k > 1$ the map $\mathcal{VL}_k$ is functorial over $\mathbf{PMet}_{inj}$ but not all of $\mathbf{PMet}$ since a non-injective map may split a k-vertex-linked subgraph [11].

Now given a finite pseudometric space (X, d_X) and $k \in \mathbb{N}$, suppose $d'_X(x_i, x_j)$ is the smallest δ such that there exists a $\min(|X|, k)$-vertex-connected subgraph of the δ-Vietoris-Rips complex of (X, d_X) that contains both x_i and x_j. Then the $\mathbf{PMet}_{bij}$-manifold learning functor $MDS \circ \mathcal{VL}_k$ maps (X, d_X) to the k-**Vertex Connected Scaling** embedding optimization problem $(|X|, m, l)$ where $l(A) = \sum_{i,j \in 1...|X|} (d'_X(x_i, x_j) - \|A_i - A_j\|)^2$. Note that unlike $MDS \circ \mathcal{L}_k$, for $k > 1$ the map $MDS \circ \mathcal{VL}_k$ is not functorial over all of $\mathbf{PMet}_{sur}$ since $\mathcal{VL}_k$ is not functorial over $\mathbf{PMet}_{sur}$.

2.3.5 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:

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

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:

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

Proposition 6. *There exists a hierarchical* $\mathbf{PMet}_{isom}$*-clustering functor FuzzySimplex that decomposes into the composition of four functors that:*

1. *build a local uber-metric space around each point in X;*
2. *convert each local uber-metric space to a fuzzy simplicial complex;*
3. *take a fuzzy union of these fuzzy simplicial complexes;*
4. *convert the resulting fuzzy simplicial complex to a fuzzy non-nested flag cover.*

(Proof in Appendix A.6)

Proposition 7. *There exists a functor* $FCE : \mathbf{FCov}_{bij} \to \mathbf{FLoss}$ *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 A.7)*

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 $\mathbf{PMet}_{bij}$.

3 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 $G(\delta) \to F(\delta + \epsilon)$ [9, 10]. 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 **FCov** or **FLoss** whose domain is $(0, 1]^{\mathrm{op}}$ 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 8. *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 A.8)*

Proposition 8 is similar in spirit to previous results that use the Gromov-Hausdorff distance between metric spaces to bound the bottleneck or homotopy interleaving distances between their corresponding Vietoris-Rips complexes [10, 7, 19, 5]. 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 prove the following:

Proposition 9. *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 A.9)

For a very simple example, consider Multidimensional Scaling without dimensionality reduction. In this case $M = MDS \circ \mathcal{ML}$ and $(X,d_X),(Y,d_Y)$ are each finite ordered n-element subspaces of $\mathbb{R}^m$ with Euclidean distance. If we write the vectors in X and Y as matrices $A_X,A_Y \in \mathbf{Mat}_{n,m}$, then these matrices respectively minimize $\mathbf{l}_{M(X,d_X)}$ and $\mathbf{l}_{M(Y,d_Y)}$, and $\mathbf{l}_{M(X,d_X)}(A_X) = \mathbf{l}_{M(Y,d_Y)}(A_Y) = 0$. Since the function that sends the ith element of X to the ith element of Y must be an $\inf\{2\epsilon \mid \forall_i \|A_{X_i} - A_{Y_i}\| \leq \epsilon\}$-isometry, Proposition 9 bounds the average of the squared distances between the paired rows of two matrices in terms of largest such distance.

These bounds apply to a very general class of manifold learning algorithms, including topologically unstable algorithms like IsoMap [4]. 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 $\forall_{i=1...n} d_X(X_i,Y_i) = d_Y(X_i,Y_i) = \|X_i - Y_i\| \leq \epsilon$. Then for any matrix $A_Y \in \mathbf{Mat}_{n,1}$ that minimizes $\mathbf{l}_{M(Y,d_Y)}$, Proposition 9 bounds the average squared difference between $|A_{Y_{i+1}} - A_{Y_i}|$ and $2r\sin(\frac{2\pi}{2n})$.

4 Experiments 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 (and therefore also stability by Proposition 9). For example, if we compose the *FuzzySimplex* functor from Proposition 6 with *MDS* we form the $\mathbf{PMet}_{isom}$-manifold learning functor $MDS \circ FuzzySimplex$ that maps (X,d_X) to the embedding optimization problem $(|X|,m,l)$ where $l(A) = \sum_{i,j\in 1...|X|}(-\log(\alpha_{ij}) - \|A_i - A_j\|)^2$ and α_{ij} is the strength of the fuzzy simplex that UMAP forms between x_i and x_j.

For a more illustrative example, 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).

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 2.1 to adapt the Metric Multidimensional Scaling algorithm $MDS \circ \mathcal{ML}$ (Section 2.3.1) 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 this 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 $|X| \times |X|$ matrix B to all zeros
  $\forall i, j \leq |X|$
    $B_{ij} = \inf\{\delta \mid \exists x_i = x_1, x_2, ..., x_n = x_j \in X, d_X(x_k, x_{k+1}) \leq \delta\}$
  $A \leftarrow \min_{A \in \mathbf{Mat}_{|X|,m}} \sum_{i,j \in 1...|X|} (||A_i - A_j|| - B_{ij})^2$
  return A

5 Discussion and Future Work

We have taken the first steps towards a categorical framework for manifold learning. By defining an algorithm as a functor from a category of metric spaces, we can explicitly express the kind of dataset transformations under which it is equivariant. We show that for many popular manifold learning algorithms, including Metric Multidimensional Scaling and IsoMap, the optimization objective changes in a predictable way as we modify the metric space.

The functorial perspective also suggests a new strategy for exploratory data analysis with manifold learning. Since we can decompose manifold learning algorithms into two components (clustering and

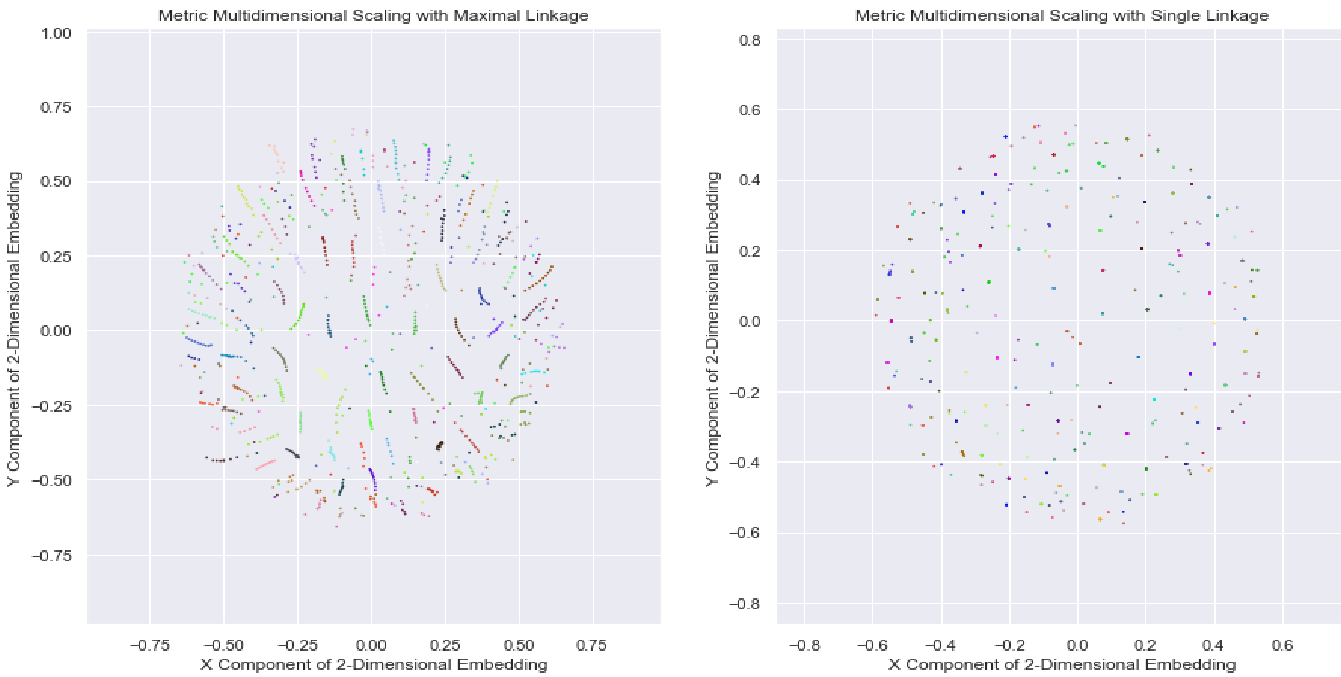

Figure 1: Embeddings of DNA sequences from the DNA recombination task with $L = 1000, N = 100, M = 10$. Each color indicates a unique DNA sequence mutation list. Note that Single Linkage Scaling ($MDS \circ \mathcal{SL}$) on the right embeds sequences in the same mutation list more closely together than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

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: Accuracy on the DNA recombination task of the Metric Multidimensional Scaling ($MDS \circ \mathcal{ML}$) and Single Linkage Scaling ($MDS \circ \mathcal{SL}$) algorithms (higher numbers are better). 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 $L = 1000$.

loss) we can examine how slight variations of the clustering algorithm affect the learned embeddings. We show in Section 2.1 that every manifold learning functor $L \circ H$ lies on a spectrum of interconnectedness between $L \circ \mathcal{ML}$ and $L \circ \mathcal{SL}$, and we can form new algorithms by moving along this spectrum. For example, we see in Section 4 that replacing the $\mathcal{ML}$ functor with $\mathcal{SL}$ in the Metric Multidimensional Scaling algorithm substantially changes the learned embeddings and improves performance on a DNA recombination task. There are also many algorithms that lie between these two options, including the k-Path Scaling and k-Vertex-Connected Scaling algorithms that we introduce in Section 2.3.

Another major benefit of expressing algorithms as functors is that functors preserve categorical properties like interleaving distance. This allows us to easily reason about the stability properties of both existing algorithms and new algorithms that we create by recombining functors. Other authors have used this strategy to prove stability properties of the homology of geometric filtered complexes [10]. In Section 3 we use this strategy to define bounds on how dataset noise affects optimization quality. In future work we hope to use these techniques to derive more powerful theorems around the resilience of other kinds of unsupervised or supervised algorithms to noise. For example, we may also be able to tighten our bounds by switching our perspective from finite metric spaces to distributions [6] or even involving categorical probability [14] to replace interleaving distance with a probabilistic analog. Due to the simplicity and flexibility of this strategy, other researchers have begun to develop more flexible characterizations of interleaving distance that we can apply in even more situations [19].

References

[1] Hervé Abdi (2007): *Metric multidimensional scaling (MDS): analyzing distance matrices*. *Encyclopedia of Measurement and Statistics. SAGE Publications.*

[2] Hervé Abdi & Lynne J Williams (2010): *Principal component analysis*. *Wiley interdisciplinary reviews: computational statistics* 2(4), pp. 433–459.

[3] Stephen Bailey (2012): *Principal Component Analysis with Noisy and/or Missing Data*. *Publications of the Astronomical Society of the Pacific* 124(919), pp. 1015–1023. Available at `http://www.jstor.org/stable/10.1086/668105`.

[4] Mukund Balasubramanian (2002): *The isomap algorithm and topological stability*. *Science* 295(5552).

[5] Andrew J Blumberg & Michael Lesnick (2017): *Universality of the homotopy interleaving distance*. *arXiv preprint arXiv:1705.01690.*

[6] Adam Brown, Omer Bobrowski, Elizabeth Munch & Bei Wang (2020): *Probabilistic convergence and stability of random mapper graphs*. *Journal of Applied and Computational Topology*, pp. 1–42.

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

[8] Gunnar Carlsson & Facundo Mémoli (2013): *Classifying clustering schemes*. *Foundations of Computational Mathematics.*

[9] Frédéric Chazal, David Cohen-Steiner, Marc Glisse, Leonidas J Guibas & Steve Y Oudot (2009): *Proximity of persistence modules and their diagrams*. In: *Proceedings of the twenty-fifth annual symposium on Computational geometry*, pp. 237–246.

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

[11] Jared Culbertson, Dan P Guralnik, Jakob Hansen & Peter F Stiller (2016): *Consistency constraints for overlapping data clustering*. *arXiv preprint arXiv:1608.04331.*

[12] Jared Culbertson, Dan P Guralnik & Peter F Stiller (2018): *Functorial hierarchical clustering with overlaps*. *Discrete Applied Mathematics* 236, pp. 108–123.

[13] Brendan Fong & David I. Spivak (2019): *An Invitation to Applied Category Theory: Seven Sketches in Compositionality*. Cambridge University Press.

[14] Tobias Fritz (2020): *A synthetic approach to Markov kernels, conditional independence and theorems on sufficient statistics*. Advances in Mathematics 370, p. 107239.

[15] Samuel Gerber, Tolga Tasdizen & Ross Whitaker (2007): *Robust non-linear dimensionality reduction using successive 1-dimensional Laplacian eigenmaps*. In: Proceedings of the 24th international conference on Machine learning, pp. 281–288.

[16] Jon M Kleinberg (2003): *An impossibility theorem for clustering*. Advances in Neural Information Processing Systems.

[17] Tom Leinster (2016): *Basic Category Theory*. Cambridge University Press.

[18] Leland McInnes, John Healy & James Melville (2018): *UMAP: Uniform manifold approximation and projection for dimension reduction*. arXiv preprint arXiv:1802.03426.

[19] Luis N Scoccola (2020): *Locally Persistent Categories And Metric Properties Of Interleaving Distances*. PhD Thesis at Western University (Ontario).

[20] Dan Shiebler (2020): *Functorial Clustering via Simplicial Complexes*. Topological Data Analysis and Beyond Workshop at NeurIPS 2020.

[21] David I Spivak (2012): *Metric realization of fuzzy simplicial sets*. Self published notes. Available at `http://math.mit.edu/~dspivak/files/metric_realization.pdf`.

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

A Appendix: Proofs

A.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, C_{X_a}) where C_{X_a} is the maximally linked sets of the relation R in which $x_1 R x_2$ if $d_X(x_1, x_2) \leq -log(a)$ [20, 11]. Consider also the function *Real* that maps the fuzzy non-nested cover (X, C_{X_a}) to the pseudometric space (X, d'_X) in which:

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

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. □

A.2 Proof of Proposition 2

Proof. By Proposition 6 in [20], 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. □

A.3 Proof of Proposition 3

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

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

A.4 Proof of Proposition 4

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

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

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

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 $MDS : \mathbf{FCov}_{sur} \to \mathbf{FLoss}$ 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 $MDS : \mathbf{FCov}_{sur} \to \mathbf{FLoss}$ 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.

□

A.5 Proof of Proposition 5

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 A.4 we have:

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

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

A.6 Proof of Proposition 6

Before we begin, we will show the following:

Proposition 10. *The category of fuzzy simplicial complexes and bijective simplicial maps* $\mathbf{FSCpx}_{bij}$ *[20] 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 6.

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): $LocalMetric_N : \mathbf{PMet}_{isom}(N) \to \mathbf{UMet}_{bij}^{\mathbf{N}}$ 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:

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

$LocalMetric_N$ 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 [20]:

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

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 $colim_N : \mathbf{FSCpx}_{bij}^{\mathbf{N}} \to \mathbf{FSCpx}_{bij}$ which sends an indexed set of fuzzy simplicial complexes in $\mathbf{FSCpx}_{bij}^{\mathbf{N}}$ to its logical fuzzy union. This functor exists by Proposition 10. In a logical fuzzy union the strength of a simplex is defined to be its maximum strength among the complexes we are adjoining* . For step 4 (convert the resulting fuzzy simplicial complex to a fuzzy non-nested flag cover), we use the functor $(Flag \circ -)$ from [20]. 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 $\mathbf{FCov}_{bij}$:

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

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.

□

A.7 Proof of Proposition 7

Proof. Define $FCE : \mathbf{FCov}_{bij} \to \mathbf{FLoss}$ to map the fuzzy non-nested cover $H : (0,1]^{op} \to \mathbf{Cov}_{bij}$ with vertex set X to $F : (0,1]^{op} \to \mathbf{Loss}$ where $F(a) = (|X|, \{c_{F(a)_{ij}}, e_{F(a)_{ij}}\})$ 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 $FCE : \mathbf{FCov}_{bij} \to \mathbf{FLoss}$ 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 $FCE(H_X)(a) = (|X|, \{c_{F(a)_{ij}}, e_{F(a)_{ij}}\})$ and $FCE(H_{X'})(a) = (|X'|, \{c'_{F(a)_{ij}}, e'_{F(a)_{ij}}\})$. 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 $FCE : \mathbf{FCov}_{bij} \to \mathbf{FLoss}$ trivially preserves the identity we can conclude that it is a functor and $FCE \circ FuzzySimplex$ is a $\mathbf{PMet}_{isom}$-manifold learning functor.

*This is different from the probabilistic simplicial complex union that the UMAP python code uses [18].

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. □

A.8 Proof of Proposition 8

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 **Loss** is a preorder, this implies that $M(X, d_X)$ and $M(Y, d_Y)$ are ϵ-interleaved. □

A.9 Proof of Proposition 9

Proof. By using Proposition 8, we see that in order to prove Proposition 9 we simply need to show that if F, G are ϵ^*-interleaved functors in **FLoss** 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:

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

□