# SIMPLICIAL REGULARIZATION

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Inspired by the fuzzy topological representation employed in UMAP (McInnes et al., 2018), we propose a regularization principle based on the preservation of the simplicial complex structure of the data. We analyze the behavior of our proposal in contrast with the common *mixup* (Zhang et al., 2018) framework on dimensionality reduction and supervised learning tasks, and show how simplicial regularization alleviates some of the shortcomings of state-of-the art methods for regularization based on data augmentation.

Figure 1: Comparison of several vicinal distributions on the MNIST dataset.

The richness of the class of neural networks as function approximators is well documented (Hornik, 1991; Cybenko, 1989; Telgarsky, 2015) and comes at the expense of having to regularize such functional families. Recent works by Zhang et al. (2018) and Verma et al. (2019) have considered the properties of regularization induced by considering convex combinations of (latent representations of) data samples. The authors argue that these techniques lead to learning "flatter" representations.

The idea of using data augmentation as a form of regularization is not new. Chapelle et al. (2000) introduced the framework of *Vicinal Risk Minimization* (VRM) in which artificial samples are generated via the application of additive Gaussian noise to the observed covariates.

In contrast, *mixup* (Zhang et al., 2018) proposes a *vicinal distribution* (Chapelle et al., 2000) via a data-agnostic augmentation procedure based on random convex combinations of elements of the dataset. For example, for a multi-class classification problem with inputs $x_i$ and one-hot targets $y_i$, *mixup* generates virtual training samples as:

$$(\tilde{x}, \tilde{y}) = \lambda(x_i, y_i) + (1 - \lambda)(x_j, y_j), \tag{1}$$

where $\lambda$ follows a distribution with support on $[0, 1]$, such as $\lambda \sim \text{Beta}(\alpha, \alpha)$, for $\alpha > 0$.

In spite of their successful empirical results, the use of the *mixup* regularization scheme is mainly justified heuristically. In this work, we study the introduction of regularization based on the preservation of topological structures present in the dataset. Our main contribution[1] is the introduction of **simplicial regularization**: we show how a regularization objective that generalizes the *mixup* framework arises naturally, when aiming to preserve the assumed simplicial complex structure of the data. Figure 1 illustrates the differences between sets of samples produced by *mixup*, VRM and our proposed simplicial regularization.

Our idea of simplicial regularization is inspired by the topology-preserving framework of UMAP (McInnes et al., 2018). At its core, this dimensionality reduction technique constructs a fuzzy topological representation of the high-dimensional dataset and then tunes the locations of the low-dimensional embedding so that the corresponding fuzzy topological representation, when measured in the low-dimensional space, matches that of the high-dimensional space.

We refer the readers to Appendix A.1 for definitions of the concepts from algebraic geometry and fuzzy sets employed below; to Appendix A.2 for an overview of the `FuzzyTop` algorithm used by UMAP to construct fuzzy topological representations; and to Appendix A.3 for a details on the `FSC-Sample` algorithm to efficiently sample from a random simplicial complex.

[1] Our code is available at: `Link available after the double-blind review period.`

### SIMPLICIAL REGULARIZATION

Consider two geometric simplicial complexes $\mathcal{K} \subset \mathbb{R}^n$ and $\mathcal{L} \subset \mathbb{R}^m$ and a function $f : \mathcal{K} \to \mathcal{L}$. We say $f$ is a simplicial map if it is "linear under convex combinations on the simplices of $\mathcal{K}$". This means we do *not* require $f$ to be a globally linear transformation, but only to commute with convex combinations of vertices spanning a simplex in $\mathcal{K}$. In other words, if $\sigma$ is a $k$-simplex in $\mathcal{K}$, we require that for every convex coefficient $k$-vector $\boldsymbol{\lambda}$:

$$f\left(\sum_{j=1}^{k} \lambda_j \sigma_j^0\right) = \sum_{j=1}^{k} \lambda_j f(\sigma_j^0) \tag{2}$$

In practice, enforcing Equation ( 2) is far from trivial. However, given a loss $l : \mathcal{L} \times \mathcal{L} \to \mathbb{R}$, we can measure how far the function $f$ is from "being a simplicial map". For this we construct a family of objectives $\mathbb{L}$, parameterized by a positive scalar $\alpha$, which encourage $f$ to **preserve the simplicial structure of** $\mathcal{K}$ when mapping to $\mathcal{L}$:

$$\mathbb{L}(f, \mathcal{K}, \alpha) = \mathbb{E}_{\sigma \sim \mathcal{K}} \mathbb{E}_{\boldsymbol{\lambda} \sim \mathrm{Dir}(\dim(\sigma), \alpha)}\ l\left(f\left(\sum_{j=1}^{\dim(\sigma)} \lambda_j \sigma_j^0\right), \sum_{j=1}^{\dim(\sigma)} \lambda_j f(\sigma_j^0)\right). \tag{3}$$

In the case $\alpha = 1$, this corresponds to a uniform distribution on the standard $(k-1)$-simplex. When $\alpha \to 0$, the distribution is very peaked at the corners of the simplex, while for $\alpha > 1$, the distribution becomes more dense towards the center of the simplex. In other words, $\alpha$ inversely controls the level of interpolation which takes place.

**Algorithm 1:** Training with Simplicial Regularization.

**Data:** Dataset $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{N}$, maximum iterations $i_{\max}$, learning rate $\eta$, Dirichlet interpolation schedule $\{\alpha_\tau\}_{\tau=1}^{i_{\max}}$.

Compute fuzzy simplicial complex $\mathcal{K} = \texttt{FuzzTop}(\mathcal{D})$
Initialize predictor $f : \mathcal{X} \to \mathcal{Y}$
**for** $\tau = 1, \ldots, i_{\max}$ **do**
  Sample $\sigma \sim \mathcal{K}$ (using $\texttt{FSC-Sample}(\mathcal{K})$), and $\boldsymbol{\lambda} \sim \mathrm{Dir}(\dim(\sigma), \alpha_\tau)$
  Compute Monte-Carlo estimator $\hat{\mathbb{L}} = l\left(f\left(\sum_{j=1}^{\dim(\sigma)} \lambda_j \sigma_j^0\right), \sum_{j=1}^{\dim(\sigma)} \lambda_j f(\sigma_j^0)\right)$
  Perform optimization step on $\hat{\mathbb{L}}$ with learning rate $\eta$
**end**

We highlight several aspects about Algorithm 1: (i) $\mathcal{K}$ only needs to be computed once per dataset and can be stored and re-used for multiple training tasks. Although *mixup* does not require this computation, usually this cost is significantly lower than that of the subsequent training pipeline. (ii) Since $\mathcal{K}$ contains its vertices as 0-simplices, the setting $\alpha \to 0$ corresponds to standard empirical risk minimization. (iii) Simplicial regularization naturally incorporates variations of different dimensions in the data via its automatic inclusion of high dimensional simplices. (iv) It is possible to set up a schedule $\{\alpha_\tau\}$ for the Dirichlet interpolation coefficient, such that $\alpha_\tau$ increases with $\tau$ during training. This schedule ensures that the map $f$ performs well on the observed vertices at an early stage, and then is progressively regularized during posterior iterations.

Contrary to previous works, our proposed vicinal distribution is not a random, additive perturbation of the observed data, or an interpolation between random, unrelated data points. Rather, we start with a carefully constructed fuzzy topological representation of the dataset and regularize the predictor towards acting as a simplicial map. The *mixup* framework can be considered a special case of simplicial regularization where we employ a dense 1-skeleton with a constant membership function.

### EXPERIMENTS

In this section, we compare the behavior of simplicial regularization and *mixup* on dimensionality reduction and supervised learning tasks. We concentrate on methods based on interpolations of data points and thus, do not compare with VRM. We argue that the introduction of full-rank isotropic Gaussian noise in the VRM framework goes against the assumption of low-dimensional structure embodied by the *manifold hypothesis* (Bengio et al., 2013), which is commonplace in the representation learning literature.

**Dimensionality Reduction.** We generated a synthetic random point cloud $X$ of size 100 in $\mathbb{R}^{20}$. We endow this point cloud with a simplicial complex structure $\mathcal{K}$ via the `FuzzTop` algorithm, and produced a 2-dimensional projection $Y$ using UMAP, illustrated in Figure 2. We then trained a 2-hidden-layer neural network $f$ of 100 units per layer to approximate embedding behavior of UMAP. That is, we want $f(x_i) \approx y_i$, where the discrepancy is measured by $l = ||\cdot||_2$. We compared the performance of a network trained using the vicinal distributions proposed by *mixup* and simplicial regularization, with a simplicial complex structure given by $\mathcal{K}$. For both methods, we used an increasing schedule for the regularization parameter $\alpha \in [10^{-4}, 1]$.

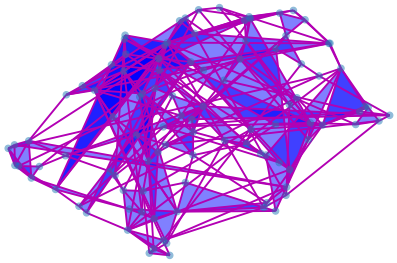

Figure 2: 2-dimensional UMAP projection of a random simplicial complex originally embedded in $\mathbb{R}^{20}$.

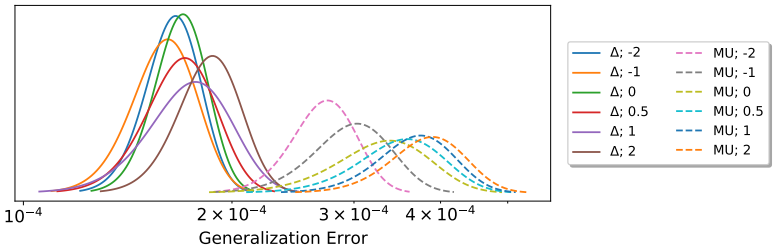


Figure 3: Effect regularization for dimensionality reduction on a synthetic simplicial complex. Numbers correspond to strength of evaluation interpolation $\alpha'$.

We evaluated the generalization performance of the networks with respect to different interpolation strengths $\alpha'$. Note that this $\alpha'$ is different from the schedule used during training, and measures the *hardness* of the task. The results are presented in Figure 3 and correspond to the distribution of 30 rounds of evaluation. As expected, we observe that the models trained using simplicial regularization perform significantly[2] better than those trained using *mixup* when the evaluation task is highly dependent on the simplicial complex structure of the data. In other words, when the test distribution involves data that is closely related to *local* interpolations of the training data points (such as tangent manifold structure (Rifai et al., 2011)), training with methods agnostic to the topological structure of the data, like *mixup*, can lead to sub-optimal performance.

**Supervised Learning.** Earlier we illustrated the qualitative differences between the inputs produced by *mixup*, VRM and simplicial regularization. In this section, we consider a digit classification task and expand our analysis of the perturbations on the data distribution induced by these techniques and the consequences at the input and prediction levels.

Figure 4: Inputs sampled from the vicinal distributions proposed by the simplicial and mixup regularization schemes with $\alpha = 1$. The variations in each column are due to sampling different convex combinations of the datapoints. The last column of the left figure corresponds to a 0-simplex.

[2] $p$-values $\approx 10^{-30}$ for a $t$-test regarding equality of means between both systems.

Figure 4 displays samples generated by interpolation of observed data points under the simplicial regularization and *mixup* frameworks on the MNIST dataset (LeCun et al., 1998). We argue that the inputs produced by simplicial regularization correspond to *plausible, local* variations of the data at the observed samples, in line with the tangent space approximation of the underlying data manifold.

This is in stark contrast with the samples produced by *mixup*, for which the unrelatedness of the samples being combined leads to inputs which can't reasonably be considered likely variations of the data distribution. It is possible to consider the use *mixup* with low $\alpha$. However, this approach towards avoiding highly improbable inputs leads to a vicinal distribution with *low interpolation* and thus, extremely close to the original data distribution, thus yielding un-regularized training. The low $\alpha$ regime does not address the main issue of *mixup*: its use of structure-agnostic interpolations.

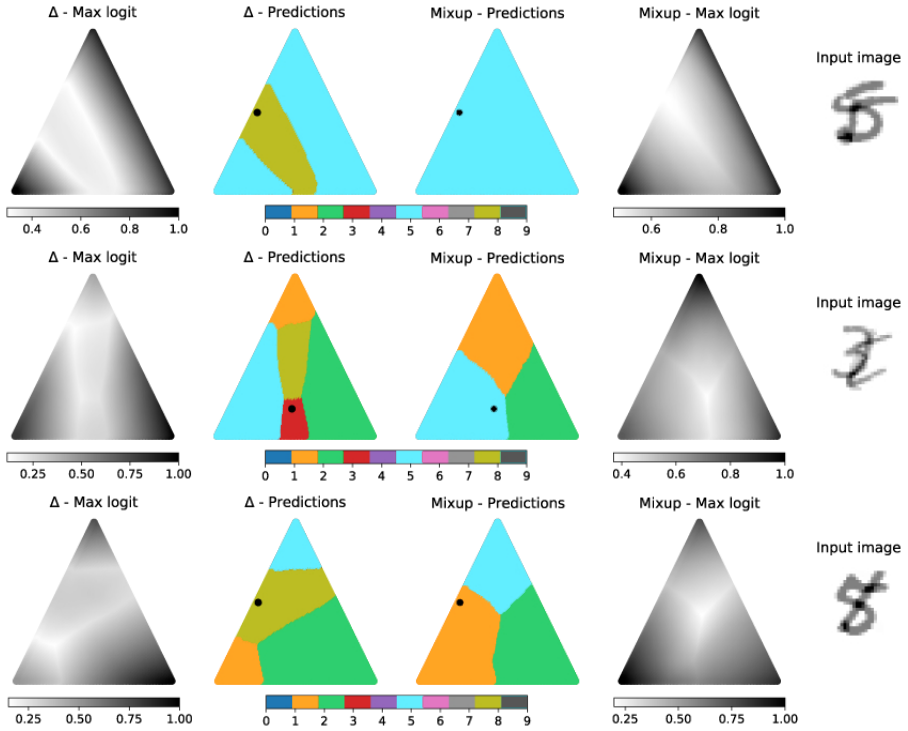


Figure 5: Behavior of classifiers on interpolation over randomly selected data points.

The *mixup* sampling scheme is also problematic regarding its proposed interpolated targets. The selection of random, unrelated, potentially very separated, data-points opens the possibility that some sections of the interpolation segment belong to regions which could be classified under a label which does not coincide with those of the endpoints. The target proposed by *mixup* is oblivious to this observation and encourages the classifier to *only* produce outputs matching the classes of the sampled endpoints.

We trained two neural networks to classify MNIST digits using *mixup* and simplicial regularization with (weighted) cross-entropy loss. We then sampled random triples of points (as in the *mixup* setting) and plotted the predictions on a grid of possible interpolations between these datapoints. These results are presented in Figure 5. Both networks reached high predictive accuracy and their predictions coincide on the corners of the simplices, yet they exhibit dissimilar behavior on interpolations.

Note how the *mixup* framework leads to classifiers which (frequently) only produce predictions which agree with the endpoint classes, even though a human observer could assign a different label. On the other hand, the local nature of the sampling in simplicial regularization allows for greater flexibility in the predictions of the model, as the training objective never aims to prescribe behavior on distant regions based solely on local information.

We hope these explorations will foster further research into the benefits of topological methods in the context of model regularization and data augmentation.

## REFERENCES

Y. Bengio, A. Courville, and P. Vincent. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1798–1828, 2013.

O. Chapelle, J. Weston, L. Bottou, and V. Vapnik. Vicinal Risk Minimization. In *NeurIPS*, 2000.

G. Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems*, 1989.

K. Hornik. Approximation capabilities of multilayer feedforward networks. *Neural networks*, 4(2): 251–257, 1991.

Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

L. McInnes, J. Healy, and J. Melville. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv preprint 1802.03426*, 2018.

J.R. Munkres. *Elements Of Algebraic Topology*. Westview Press, 1995.

S. Rifai, Y. N. Dauphin, P. Vincent, Y. Bengio, and X. Muller. The Manifold Tangent Classifier. In *NeurIPS*, 2011.

M. Telgarsky. Representation Benefits of Deep Feedforward Networks. *arXiv preprint 1509.08101*, 2015.

V. Verma, A. Lamb, C. Beckham, A. Courville, I. Mitliagkas, and Y. Bengio. Manifold Mixup: Better Representations by Interpolating Hidden States. In *ICML*, 2019.

H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz. *mixup*: Beyond Empirical Risk Minimization. In *ICLR*, 2018.

## A APPENDIX

### A.1 DEFINITIONS AND NOTATION

For completeness, we provide definitions of several topological notions used in this manuscript. While many of these definitions can be formulated in category-theoretical terms, we favour the purely geometric language for the sake of accessibility to the general machine learning community. The definitions below follow closely the work of Munkres (1995).

A $\mathcal{K}$**-simplex** $\sigma$ in $\mathbb{R}^n$ is the convex set spanned by $k+1$ linearly independent points $\{x_0, \ldots, x_k\} \subset \mathbb{R}^n$. The points $x_i$ are called *vertices*, and the convex set spanned by any non-empty subset of these vertices is called a *face* of the $\mathcal{K}$-simplex. The set of $l$-simplices contained in $\sigma$ is denoted $\sigma^l$. For example, the set of vertices of $\sigma$ is denoted $\sigma^0$ and $\sigma^k$ only contains $\sigma$ itself. For $k < l$, $\sigma^l$ is empty.

The **standard** geometric $\mathcal{K}$-simplex, denoted $\Delta^k$, is the convex hull of the canonical basis of $\mathbb{R}^{k+1}$.

A **simplicial complex** $\mathcal{K} \subset \mathbb{R}^n$ is a collection of simplices in $\mathbb{R}^n$, possibly of varying dimensions, such that: (i) every face of a simplex of $\mathcal{K}$ is also in $\mathcal{K}$, and (ii) the intersection of any two simplices of $\mathcal{K}$ is a face of each of them.

Intuitively, we can think of a simplicial complex $\mathcal{K}$ as made up of copies of standard simplices of several dimensions, glued together among some common faces. We can organize the relevant information about a simplicial complex into the different skeleta $\mathcal{K}^k$, for $k = 0, 1, \ldots$, so that $\mathcal{K}^k$ is the set of all $\mathcal{K}$-simplices of $\mathcal{K}$.

The structure-preserving functions between simplicial complexes are called **simplicial maps**. Let $\mathcal{K}$ and $\mathcal{L}$ be geometric simplicial complexes. A simplicial map $f : \mathcal{K} \to \mathcal{L}$ is a function which acts linearly on convex combinations over the simplices of $\mathcal{K}$.

Algebraically, consider an arbitrary $m$-simplex $\sigma \subset \mathcal{K}$ and a point $x \in \sigma$. It is possible to represent $x$ via its **barycentric coordinates** over $\sigma$ as a weighted sum of the vertices $\sigma^0$: there exists a

probability $m$-vector $\boldsymbol{t}^\sigma$, such that $x = \sum_{j=1}^m t_j^\sigma \sigma_j^0$. We say that $f$ is a simplicial map if all $x$ and for all $\sigma$ containing $x$, we have that:

$$f(x) = f\left(\sum_{j=1}^m t_j^\sigma \sigma_j^0\right) = \sum_{j=1}^m t_j^\sigma f(\sigma_j^0). \tag{4}$$

Note that a simplicial map $f$ is uniquely determined by a function $f_0 : \mathcal{K}^0 \to \mathcal{L}^0$ operating over the vertices, and its extension by convex interpolation on each simplex in $\mathcal{K}$.

Let $\alpha > 0$. We denote by $\mathrm{Dir}(k, \alpha)$ the **symmetric Dirichlet distribution** with density over the standard $k-1$ simplex given by $\frac{\Gamma(\alpha k)}{\Gamma(\alpha)^k} \prod_{i=1}^k x_i^{\alpha-1}$.

A **fuzzy set** is set $S$ enriched with a membership function $\mu : S \to [0,1]$. Given fuzzy sets $(S, \mu)$ and $(T, \nu)$ a morphism of between them is a function $f : S \to T$ such that for all $s \in S$, $\mu(s) \leq \nu(f(s))$.

Recalling the notion of a simplicial complex as a collection of simplices, we define a **fuzzy simplicial complex** as a simplicial complex $\mathcal{K}$ endowed with a membership function $\mu_\mathcal{K} : \mathcal{K} \to [0,1]$, such that for all $\sigma \subset \sigma'$, $\mu_\mathcal{K}(\sigma) \geq \mu_\mathcal{K}(\sigma')$. In other words, the membership of a simplex is upper bounded by the lowest membership of all its constituent lower-order simplices.

A **strong negator** is a monotonous decreasing, involutive function $\neg : [0,1] \to [0,1]$ such that $\neg 0 = 1$ and $\neg 1 = 0$. For the remainder, we restrict our attention to the negator $\neg(a) = 1 - a$.

A **t-norm** is a symmetric function $\top : [0,1]^2 \to [0,1]$ satisfying for all $a, b \in [0,1]$:

- $\top(a, b) \leq \top(c, d)$ whenever $a \leq c$ and $b \leq d$,
- $\top(a, \top(b, c)) = \top(\top(a, b), c)$, and
- $\top(1, a) = a$.

Given a t-norm $\top$, its complementary **t-conorm** under a negator $\neg$ is defined by $\perp(a, b) = \neg\top(\neg a, \neg b)$. A triplet $(\top, \perp, \neg)$ where $\top$ is a t-norm, $\perp$ is a t-conorm, and $\neg$ is a strong negator is called a **De Morgan triplet** if for all $a, b \in [0,1]$ one has that $\neg\perp(a, b) = \top(\neg a, \neg b)$.

The most common example of a De Morgan triplet is the one formed by $\top_{\text{prod}}(a, b) = ab$, $\perp_{\text{sum}} = a + b - ab$ and $\neg(a) = 1 - a$. Note how the t-norm and t-conorm express the probability of intersection and union of independent events. Another important example arises by taking $\top_{\text{min}}(a, b) = \min(a, b)$, $\perp_{\text{max}} = \max(a, b)$.

A De Morgan triplet $(\top, \perp, \neg)$ structure allows us to define operations between sets, or rather, between membership functions $\mu$ and $\nu$ on a common universal set $U$. The **complement** of $\mu$ given by the composite function $\neg \circ \mu$. We define their **union** and **intersection** as the functions $\tau_{\mu\cup\nu} = \perp \circ (\mu, \nu)$ and $\tau_{\mu\cap\nu} = \top \circ (\mu, \nu)$, respectively.

### A.2 Fuzzy representation of a dataset

The main building block of the UMAP dimensionality reduction technique is the construction of a carefully designed fuzzy topological representation of a given dataset $\mathcal{D} \subset \mathbb{R}^n$. In practice, this representation is stored as a weighted graph with restricted neighborhood size. Next, we present a brief description of Algorithm 2. The steps involved in this construction are illustrated in Fig. 6. For more details, please consult sections 2.2 and 3.1 of McInnes et al. (2018).

**Algorithm 2:** `FuzzyTop`

**Data:** Dataset $\mathcal{D} = \{x_i\}_{i=1}^N \subset \mathbb{R}^n$, number of nearest neighbors $\tilde{n}$.
**Result:** Fuzzy topological representation of $\mathcal{D}$ given by $\mathcal{K}_\mathcal{D}$.

**for** $i = 1, \ldots, N$ **do**
  $\mathcal{N}_i =$ indices of $\tilde{n}$-nearest neighbors of $x_i$;
  Compute $(\mathcal{D}, d_i) \in$ **FinEPMet** ; `// Build FinEPMet` $d_i$ `around` $i$
  $\mathcal{K}_i = \mathrm{FinSing}(\mathcal{D}, d_i) \in$ **sFuz** ; `// Convert` $d_i$ `into FSC`
**end**
$\mathcal{K}_\mathcal{D} = \perp_{i=1}^N \mathcal{K}_i$ ; `// Combine all FSCs`

We begin with a dataset $\mathcal{D}$ of $N$ points embedded in $\mathbb{R}^n$ and an arbitrary metric $d$. For each data point $i$, we construct its set of $\tilde{n}$ nearest neighbors $\mathcal{N}_i$ and record the distance to its closest neighbor in $\rho_i = \min_{j \in \mathcal{N}_i} d(x_i, x_j)$. We define a finite, extended, pseudo-metric space $d_i$ around $x_i$ as:

$$d_i(x_i, x_j) = \begin{cases} d(x_i, x_j) - \rho_i, & \text{for } j \in \mathcal{N}_i \\ \infty & \text{else.} \end{cases} \tag{5}$$

We define $\mu_{i,j} = \exp\left(-\frac{d_i(x_i, x_j)}{\sigma_i}\right)$ for $j \in \mathcal{N}_i$, and select $\sigma_i > 0$ in such a way that $\sum_{j \in \mathcal{N}_i} \mu_{i,j} = \log \tilde{n}$. These ingredients allow us to define a fuzzy simplicial complex $\mathcal{K}_i$ whose fuzzy 1-skeleton is given by the membership function $\mu_{i,(\cdot)}$. Note how all points in $\mathcal{D}$ outside of the neighborhood $\mathcal{N}_i$ have membership zero.

Finally, this process is repeated for all data points in $\mathcal{D}$ and the individual simplicial complexes are joined via a pre-determined t-conorm $\perp$.

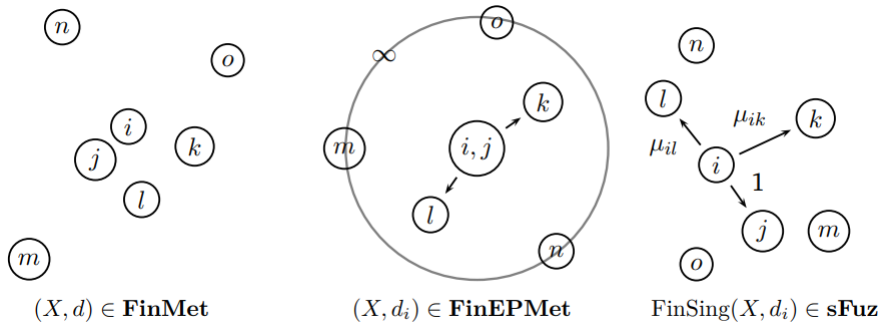


Figure 6: Visual representation of the fuzzy simplicial set construction around datapoint $i$ within the UMAP framework. In this example $\tilde{n} = 3$.

### A.3 Simplicial complex extension and sampling

Algorithm 2 and its output are manageable for large-scale datasets due to the restricted neighborhood size and the use of a 1-skeleton representation. Explicitly storing all simplices of higher dimensions quickly becomes a computational bottleneck due to the combinatorial number of possibilities. In this section we present a construction that extends a fuzzy 1-skeleton into a fuzzy simplicial complex, and an efficient algorithm to sample from it, while avoiding the explicit storage of memberships for high-dimensional simplices.

**Algorithm 3:** `FSC-Sample`

**Data:** 1-skeleton $\mu$, neighborhoods $\mathcal{N}_i$, t-norm $\top$ (e.g. $\top_{\text{min}}$)
**Result:** A random simplex sampled from the fuzzy simplicial complex extended from $\mu$.

Sample $u \in [1, \ldots, N]$ uniformly
$\mathcal{N} = \mathcal{N}_u \cup \blacksquare$
$\sigma = [u]$
**while** $\mathcal{N} \neq \{\blacksquare\}$ **do**
  **foreach** $t \in \mathcal{N}$ **do** $\nu_t = \top_{s \in \sigma} \mu_{s,t}$;
  $\nu_{\blacksquare} = 1$
  Sample $u'$ without replacement from $\mathcal{N}$ with probabilities proportional to $\nu$
  **if** $u' = \blacksquare$ **then**
    Return $\sigma$
  **else**
    Append $u'$ to $\sigma$
  **end**
  $\mathcal{N} = (\mathcal{N} \cap \mathcal{N}_{u'}) \cup \blacksquare$
**end**

Consider a (sparse) symmetric $N \times N$ matrix $\mu$ representing a fuzzy 1-skeleton and denote by $\mathcal{N}_i$ the non-zero entries in row $i$. It is easy to see that under a t-norm $\top$, the recursive membership:

$$\mu_{\mathcal{K}}(\sigma) = \top_{\psi \in \sigma} \mu_{\mathcal{K}}(\psi) \quad \text{for } \sigma \in \mathcal{K}^k,\ k > 2 \tag{6}$$

is a valid extension of $\mu$ onto a fuzzy simplicial complex. The base cases for these recursions are the 2-simplex memberships $\mu_{\mathcal{K}}([i, j]) = \mu_{i,j}$. Trivially, this construction implies that a simplex $\sigma$ has non-zero membership if *all* its constituent sub-simplices have non-zero membership.

Note that Algorithm 3 builds the required high-dimensional simplices on the fly, and is guaranteed to terminate in at most $\tilde{n}$ iterations of the outermost cycle. The sampling of multiple simplices for the mini-batch setting can be easily parallelized.