On the Impact of Topological Regularization on Geometrical and Topological Alignment in Autoencoders: An Empirical Study

Samuel Graepler
Faculty of Mathematics and Computer Science
Leipzig University
Leipzig, Germany
samuel.graepler@outlook.de

Diaaeldin Taha
Max Planck Institute for Mathematics in the Sciences
Leipzig, Germany
diaaeldin.taha@mis.mpg.de

Nico Scherf
Max Planck Institute for Human Cognitive and Brain Sciences
Leipzig, Germany

Anna Wienhard
Max Planck Institute for Mathematics in the Sciences
Leipzig, Germany

Abstract

We present a comparative empirical study on the impact of topological regularization on autoencoders (AEs) and variational autoencoders (VAEs) across six synthetic datasets with known topology and curvature. Particularly, we probe the alignment of the topology and geometry of the dimensionality-reduced latent representation with that of the data. To quantify geometrical alignment, we estimate the mean extrinsic curvature of the latent embedding by fitting local quadrics. We find that topological regularization can significantly improve the geometrical alignment of latent and data, even when the training objective emphasizes topological alignment alone, without regard for reconstruction quality.

1 Introduction

Learning meaningful representations is central to understanding and processing high-dimensional data. The *manifold hypothesis* [Bengio et al., 2013] states that real-world data often lies near a low-dimensional manifold, a view supported in computer vision [Carlsson et al., 2008], neuroscience [Chaudhuri et al., 2019], and machine learning [Naitzat et al., 2020]. From a *manifold learning* perspective, an ideal representation respects both topology and geometry: local neighborhoods, distances, and angles are preserved while enabling operations like interpolation and distance computations [Lee et al., 2022, Hauberg, 2019].

However, in practice, most methods fail to satisfy these properties reliably. State-of-the-art manifold learning methods like t-SNE and UMAP struggle to represent global structure faithfully and frequently

Preprint.

do not preserve topology [Moor et al., 2020, Nazari et al., 2023]. In geometric deep learning, autoencoders provide a popular tool to face this task. A standard autoencoder (AE) can be considered an embedding method that learns a dimensionality-reduced latent representation of the data in the latent space $\mathcal{Z}$. A probabilistic variant is the variational autoencoder (VAE) [Kingma and Welling, 2014], which instead learns a joint probability distribution over the input and latent space. For each input x, the encoder produces a distribution $\mathrm{enc}(x) = q(z|x)$ on $\mathcal{Z}$, such that a sample $z^* \sim q(z \mid x)$ can be decoded with a high probability to $\mathrm{dec}(z^*) \approx x$. Thus, every forward pass of the data yields a probabilistic embedding in the latent space. Although various extensions alter the latent geometry, the standard choice is $\mathcal{Z} = \mathbb{R}^d$, a flat Euclidean space.

In this work We present preliminary results from an empirical comparison of deterministic AEs and Gaussian VAEs with Euclidean latent spaces. We investigate their ability to learn latent representations that align topologically and geometrically with the data, using six synthetic datasets lying near one and two-dimensional embedded manifolds. Topological alignment is assessed visually, while geometry preservation is assessed through estimates of the *mean extrinsic curvature*, a measure of how the manifold bends within the ambient space. We estimate curvature directly from the latent embedding via a local quadric fit, similar to Gilbert and O'Neill [2025], Yang and Lee [1999]. In particular, we investigate the impact of *topological regularization* [Moor et al., 2020] on topological and geometric representational alignment.

2 Models and training

2.1 Datasets and architectures

To obtain controlled ground-truth data, we construct six synthetic datasets by embedding smoothly deformed one- and two-dimensional manifolds into $\mathcal{X} = \mathbb{R}^{10}$. Each manifold is homeomorphic to either a circle $\mathcal{S}^1$, a sphere $\mathcal{S}^2$, or a torus $\mathcal{T}^2$. For every topology, we generate two variants. In the *low*-deformation datasets ($\mathrm{Circle_{low}}, \mathrm{Sphere_{low}}, \mathrm{Torus_{low}}$, adapted from Acosta et al. [2023]), the d-dimensional manifold remains in a $(d+1)$-dimensional subspace of $\mathbb{R}^{10}$. In the *high*-deformation datasets, the manifold is bent across more ambient dimensions (9 for the $\mathrm{Circle_{high}}$, 5 for $\mathrm{Sphere_{high}}$ and $\mathrm{Torus_{high}}$). To all datasets, we finally apply random rotations and small Gaussian noise.

We examine standard deterministic AEs and Gaussian VAEs with Euclidean latent space. Depending on the examined dataset, we set $\mathcal{Z} = \mathbb{R}^{d+1}$, where d is its intrinsic manifold dimension (e.g. $d = 1$ for $\mathrm{Circle_{low}}$).

2.2 Topological and geometrical evaluation

We assess the topological alignment of data and latent embedding by visually inspecting the latent space, focusing on the topological signature, i.e. the number of connected components, holes and voids, compared to the ground truth of the respective dataset. This is possible since all datasets lie close to one- and two-dimensional manifolds and the latent spaces are constrained to be $\mathbb{R}^2$ or $\mathbb{R}^3$.

To evaluate the geometric alignment between the latent embedding $\{z_i\}_{i=1}^N \subset \mathcal{Z}$ and the original data $\{x_i\}_{i=1}^N \subset \mathcal{X}$, we estimate the local curvature of both manifolds. For each point in the latent space, we identify its k nearest neighbors and apply PCA to fit a quadric that locally approximates the underlying manifold structure. The mean extrinsic curvature is then computed from the Hessian of the fitted quadric, following the methods of Gilbert and O'Neill [2025], Yang and Lee [1999]. We measure the geometrical similarity of data and latent representation using both absolute and relative error metrics on curvature estimates κ:

$$\mathrm{MSE} = \frac{1}{N}\sum_i (\kappa(x_i) - \kappa(z_i))^2, \quad \mathrm{SMAPE} = \frac{1}{N}\sum_i \frac{|\kappa(x_i) - \kappa(z_i)|}{|\kappa(x_i)| + |\kappa(z_i)|}.$$

2.3 Intervention and Training

Persistent homology provides a way to estimate the topology of the underlying manifold of a point cloud. We can extract topological properties by tracking the birth and death of connected components, loops, and higher-dimensional cycles across a series of simplicial complexes with increasing connectivity. The Topological Autoencoder Moor et al. [2020] integrates this into training via a differentiable regularization term $\mathcal{L}_t$, aligning latent and input topology by matching lengths of *topologically relevant* edges, which give birth and death to topological features. Let $A^{\mathcal{X}}$, $A^{\mathcal{Z}}$ be the pairwise distance matrices and $\pi^{\mathcal{X}}$, $\pi^{\mathcal{Z}}$ the corresponding persistence pairings, and denote $A[\pi]$ as the extraction of the lengths of topologically relevant edges. Then

$$\mathcal{L}_t = \frac{1}{2}\|A^{\mathcal{X}}[\pi^{\mathcal{X}}] - A^{\mathcal{Z}}[\pi^{\mathcal{X}}]\|^2 + \frac{1}{2}\|A^{\mathcal{Z}}[\pi^{\mathcal{Z}}] - A^{\mathcal{X}}[\pi^{\mathcal{Z}}]\|^2.$$

To encourage topological alignment, $\mathcal{L}_t$ is added to the training objective:

$$\mathcal{L} = \alpha\mathcal{L}_{\text{recon}} + \gamma\mathcal{L}_{\text{t}} \quad \text{(for AEs)}, \qquad \mathcal{L} = \alpha\mathcal{L}_{\text{recon}} + \beta\mathcal{L}_{\text{KL}} + \gamma\mathcal{L}_{\text{t}} \quad \text{(for VAEs)}.$$

We fix a hyperparameter $\dim_t$, defining the highest topological feature dimension considered (e.g. $\dim_t = 2$ captures components, loops, and voids), and train each dataset–architecture pair with low, medium, and high weights $\alpha \in \{0, 1\}$, $\beta \in \{0, 0.08, 1\}$, $\gamma \in \{0, 1, 100\}$.

The topological loss introduces additional computational overhead primarily through the persistent homology calculation performed on each mini-batch. This cost increases when higher-dimensional topological features are considered. However, consistent with the observations of Moor et al. [2020], when restricting the regularization to low-dimensional features, the use of mini-batches limits the effective size of the point cloud on which persistence is computed, so the computational cost scales primarily with the mini-batch size rather than the full dataset size.

3 Results

3.1 Autoencoders

Autoencoders trained without topological regularization show topological but not geometrical alignment between data and the latent embedding. Including topological regularization can substantially improve geometric alignment (see Figure 2 and Figure 1), even when only regularizing on topology, without including the reconstruction loss in the training objective ($\alpha = 0$), leaving the decoder unconstrained. For datasets lying in a low-dimensional subspace, topological regularization yields near-perfect embeddings. However, the effect of topological regularization depends strongly on the choice of the considered topological features ($\dim_t$), and inappropriate settings can even disrupt topological alignment. Notably, no monotonic improvement in geometrical alignment with increasing topological feature dimension was observed. This suggests that for specific choices of the topological feature dimension, regularizing only the topologically relevant edges within each mini-batch not only preserves the global topology of the dataset, but also, somewhat surprisingly, promotes local pairwise distance consistency, leading to improved geometrical alignment.

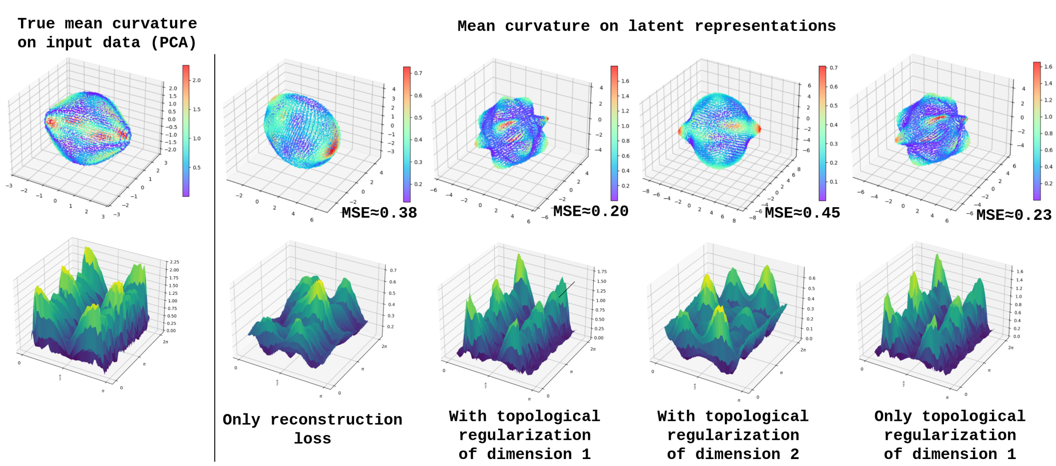

Figure 1: Impact of topological regularization on the latent geometry of an AE trained on the $\mathrm{Sphere}_{\mathrm{high}}$ dataset. The upper row shows the curvature of the input data and the learned representation as heatmap. The lower row shows the latent curvature plotted over the ground truth angles.

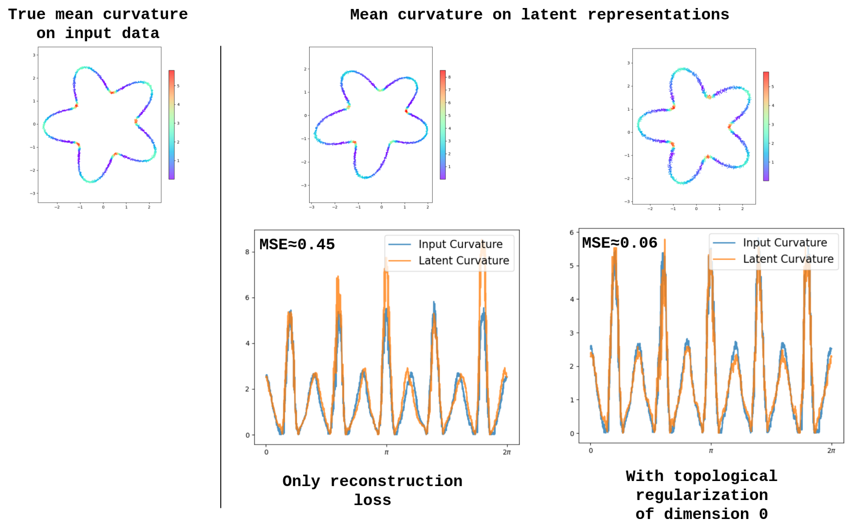

Figure 2: Impact of topological regularization on the latent geometry of an AE trained on the $\mathrm{Circle}_{\mathrm{low}}$ dataset. The upper row shows the curvature of the input data and the learned representation as heatmap. The lower row shows the latent curvature plotted against input curvature over ground truth angles.

Table 1: Mean squared error (MSE) between curvature estimates of the input data and the autoencoders' latent representations across datasets and weight settings.

α	γ	dim_t	$\mathrm{Circle}_{\mathrm{low}}$	$\mathrm{Circle}_{\mathrm{high}}$	$\mathrm{Sphere}_{\mathrm{low}}$	$\mathrm{Sphere}_{\mathrm{high}}$	$\mathrm{Torus}_{\mathrm{low}}$	$\mathrm{Torus}_{\mathrm{high}}$
1	0	–	0.45	2.7776	0.4102	0.3836	0.0329	0.8026
1	1	0	**0.0630**	10.8412	0.0504	0.3024	0.0009	**0.2754**
1	100	0	0.5671	12.6125	0.2032	0.2784	**0.0004**	<u>0.3150</u>
0	1	0	0.5587	12.8751	0.2081	0.3049	<u>0.0006</u>	0.3171
1	1	1	<u>0.0820</u>	3.6430	**0.0003**	**0.1981**	0.0158	0.7181
1	100	1	1.9507	2.7692	0.0092	0.2468	0.0203	0.9048
0	1	1	1.7502	2.6157	<u>0.0077</u>	<u>0.2301</u>	0.0184	0.8989
1	1	2	109.6815	**2.1280**	0.7751	0.4535	0.0483	0.8652
1	100	2	2342.8889	2.4720	0.8100	0.4580	0.1506	1.0173
0	1	2	64217.8867	<u>2.3382</u>	0.7440	0.4682	0.1554	1.0158

3.2 Variational Autoencoders

For VAEs without topological regularization, the KL-loss term ($\beta > 0$) pulls the latent embedding toward a Gaussian blob, preventing both topology and geometry preservation. Topological regularization can partly counteract this effect for low weights on the KL-loss term, but the latent geometry remains disrupted. When the KL term is omitted, VAEs behave similarly to AEs (see Figure 3).

4 Discussion and conclusions

We conducted an empirical study of autoencoders and variational autoencoders, evaluating their ability to preserve data topology and geometry in latent embeddings across six synthetic datasets. Local geometry was quantified via mean extrinsic curvature, estimated on the learned representation $\{z_i\}_{i=1}^N$ by fitting a local quadric.

For AEs, topological regularization $\mathcal{L}_t$ significantly improved geometric alignment between latent representations and data, even in the absence of the reconstruction objective. For VAEs, topological

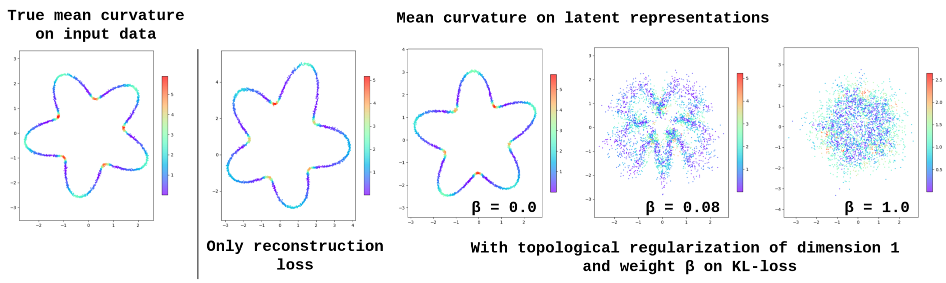

Figure 3: Impact of the KL term and topological regularization on the latent geometry of a VAE trained on the $\mathrm{Circle}_{\mathrm{low}}$ dataset. The KL-loss term ($\beta > 0$) disrupts topological and geometrical alignment.

regularization acted as a counterforce to the origin gravity created by the KL regularization term, but the resulting representations remained noisy and failed to preserve topology or geometry.

These findings suggest that (i) reliably topologically and geometrically aligned representations can potentially be learned by regulating only the pairwise distances between points, as indicated by the impact of the topological regularization term, and (ii) this may be achievable with a single neural network rather than a full autoencoder architecture, since the decoder became redundant when omitting the reconstruction objective.

4.1 Limitations and next steps

An immediate next step is to investigate whether architectures optimizing only for topological alignment without a decoder can yield reliable representations.

We evaluated local geometry using extrinsic curvature estimated via local quadrics, a naive method that restricts analysis to simple datasets. Future work should aim to broaden the scope toward more diverse and real-world datasets.

Topological alignment was only indirectly assessed. A similarity score based on persistent homology would provide a stronger measure. Future work should also examine intrinsic curvature, which is independent of ambient geometry.

Finally, repeated experiments are needed to test whether topological regularization also leads to representational alignment across runs.

References

Francisco Acosta, Sophia Sanborn, Khanh Dao Duc, Manu Madhav, and Nina Miolane. Quantifying Extrinsic Curvature in Neural Manifolds. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 610–619, Vancouver, BC, Canada, June 2023. IEEE. ISBN 979-8-3503-0249-3. doi: 10.1109/CVPRW59228.2023.00068.

Y. Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35:1798–1828, 08 2013. doi: 10.1109/TPAMI.2013.50.

Gunnar Carlsson, Tigran Ishkhanov, Vin de Silva, and Afra Zomorodian. On the local behavior of spaces of natural images. *International Journal of Computer Vision*, 76(1):1–12, 2008. ISSN 1573-1405. doi: 10.1007/s11263-007-0056-x. URL `https://doi.org/10.1007/s11263-007-0056-x`.

Rishidev Chaudhuri, Berk Gerçek, Biraj Pandey, Adrien Peyrache, and Ila Fiete. The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep. *Nature Neuroscience*, 22(9):1512–1520, 2019. ISSN 1546-1726. doi: 10.1038/s41593-019-0460-x. URL `https://doi.org/10.1038/s41593-019-0460-x`.

Anna C Gilbert and Kevin O'Neill. Ca-pca: Manifold dimension estimation, adapted for curvature. *SIAM Journal on Mathematics of Data Science*, 7(1):355–383, 2025.

Søren Hauberg. Only Bayes should learn a manifold (on the estimation of differential geometric structure from data), 2019. URL http://arxiv.org/abs/1806.04994.

Diederik P. Kingma and Max Welling. Auto-Encoding Variational Bayes. In *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*, 2014.

Yonghyeon Lee, Sangwoong Yoon, MinJun Son, and Frank C. Park. Regularized autoencoders for isometric representation learning. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=mQxt8l7JL04.

Michael Moor, Max Horn, Bastian Rieck, and Karsten Borgwardt. Topological autoencoders. In Hal Daumé III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 7045–7054. PMLR, 7 2020. URL https://proceedings.mlr.press/v119/moor20a.html.

Gregory Naitzat, Andrey Zhitnikov, and Lek-Heng Lim. Topology of deep neural networks. *Journal of Machine Learning Research*, 21(184):1–40, 2020. URL http://jmlr.org/papers/v21/20-345.html.

Philipp Nazari, Sebastian Damrich, and Fred A. Hamprecht. Geometric autoencoders: what you see is what you decode. In *Proceedings of the 40th International Conference on Machine Learning*, ICML'23. JMLR.org, 2023.

M. Yang and E. Lee. Segmentation of measured point data using a parametric quadric surface approximation. *Computer-Aided Design*, 31(7):449–457, 1999. ISSN 0010-4485. doi: https://doi.org/10.1016/S0010-4485(99)00042-1. URL https://www.sciencedirect.com/science/article/pii/S0010448599000421.