Topo Sampler: A Topology Constrained Noise Sampling for GANs

Adrish Dey *
Department of Computer Science
Netaji Subhash Engineering College
Kolkata, West Bengal
adrishd.cse2017@nsec.ac.in

Sayantan Das *
Department of Computer Science
Institute of Engineering and Management
Kolkata, West Bengal
sdas.codes@gmail.com

Abstract

This work studies disconnected manifold learning in generative adversarial networks in the light of point set topology and persistent homology. Under this formalism, the topological similarity of latent space in generative models with the underlying manifold of data distribution facilitates better generalization. To achieve this, we introduce a topology-constrained noise sampler, responsible for mapping the samples from Gaussian spheres to a latent embedding space, which in turn is constrained to be topologically similar to the manifold underlying the data distribution. We study the effectiveness of this method in GANs for learning disconnected manifolds. This is an ongoing research, with the current report containing preliminary empirical experiments. The codebase for our experiments is publicly available at https://github.com/captain-pool/TopoSampler

1 Introduction

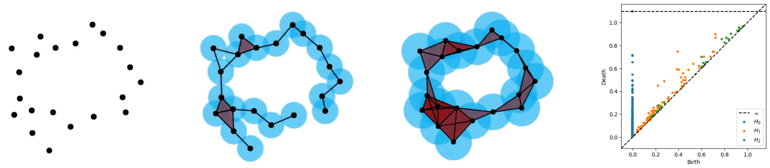
Learning distributions with disconnected support is a challenging problem in generative models, especially GANs. Initially formalized by [16], the problem of learning distributions on a disconnected support with unimodal Gaussian distribution is still an open area of research. Over the years various methods have been proposed to tackle this problem. This includes [25]'s rejection sampling of non-manifold samples, [2]'s rejection sampling in latent space, and [16]'s multi-generator approach.

In this work, we focus on the *unimodal distribution* part of the problem and exploit the topological properties of the underlying manifold of the data distribution to learn a complex initial distribution, *ideally* having the same number of connected components (or 0D holes). Finally we learn a homeomorphic transformation between the disconnected distribution obtained from the sampler, and the disconnected distribution of the data, by constraining the GAN in a novel manner. To sum up, our paper makes the following contributions:

- We make novel assertions about the guarantee of the existence of a homeomorphism in a 1-lipschitz constrained generator for disconnected support learning.
- The above guarantee enables us to study the effects of Topological Regularization of the Latent space in WGANs.
- We study a Persistent Homology based method to regularize the topological properties of the sample space of probability distributions.

*equal contributions. order chosen by a coin-toss

Topological Data Analysis and Beyond Workshop at the 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.



(a) (b) (c) (d)

Figure 1: Vietoris-Rips Complex $\mathfrak{R}_\epsilon(X)$ of a point cloud X at different scales, $\epsilon_0 < \epsilon_1 < \epsilon_2$. 1(d) shows the Persistence Diagram obtained from Vietoris-Rips Complex. The diagram is represented as a set of points of form $(\epsilon_{birth}, \epsilon_{death}) \in \mathbb{R} \times \mathbb{R} \cup \{\infty\}$. where ϵ_{birth} represents the radius of ϵ-ball at which a feature is born, and ϵ_{death} represents the radius when a feature is closed by a higher-dimensional simplex. Figures 1(a), 1(b), 1(c) taken from https://aqjaffe.github.io/VRPolygons/

2 Background

Persistent Homology (PH) [4, 7] is an algorithm for extracting topological features from data sets. In our paper we use TDA methods to create a regularized WGAN objective to learn a disconnected support. We therefore need to briefly introduce some of the prime topics in this section. The Vietoris–Rips Filtration refers to a specific form of simplicial complexes, calculated from point cloud data. A simplicial complex is a generalized graph, or, intuitively speaking, a triangulated space. A simplicial complex $\mathcal{K}$ is a collection of simplices such that

1. Every face of the simplex $\mathcal{K}$ resides in $\mathcal{K}$.
2. The intersection of any two simplices of $\mathcal{K}$ is a face of each of them.

A filtration of a simplicial complex is a sequence of nested subcomplexes. Vietoris–Rips (VR) complexes are defined on a finite set of d-dimensional points, i.e. a point cloud, with an additional distance metric, for example the Euclidean metric. Roughly speaking, the metric is required to gauge the "scale" at which the simplicial complex is being generated. Given a threshold ϵ, the Vietoris–Rips Complex of a set of points X with euclidean metric d is defined as

$$V_\epsilon(X) = \{\sigma \subseteq X | d(u, v) \leq \epsilon, \forall u \neq v \in \sigma\} \tag{1}$$

Given a VR complex, persistent homology refers to a method that associates a set of multi-scale topological features to it. Such features are stored in persistence diagrams, which constitute a multiset representation of the birth and death values (measured using the distance function used for the creation of the VR complex) of the topological features of the VR complex. Recent works such as [22] use the fact that a infinitesimal change in the underlying point cloud can be easily accommodated owing to the stable nature of Persistence diagrams to leverage a topological signature distance in their paper. We use this particular result as a basis for establishing a discrepancy score between data and latent manifolds.

3 Related Work

Previous works have evaluated Generative models to facilitate disconnected manifold learning [16, 17], paving the ground for latent surgical techniques as in [2, 25]. However, we feel inserting disconnectedness into the generator's proposal distribution is not the right perspective. Instead, in our work, we study the simplicial homology of the samples from latent manifold and leverage topological descriptors to learn a Neural Sampler to generate a latent distribution whose sample space manifold is homologically similar to the original data manifold. This is an unprecedented direction towards learning distributions with disconnected supports and arguably the first in this domain.However this should not be confused with Horak et al. [14], which is an evaluation technique motivated by TDA. [13] showed empirically the generalising effects of topological regularisation between the samples from desired probability distribution and the high dimensional features of the linear classifier. This is also one of the preliminary works which uses a differentiable PH calculation method. Also from the Optimal Tranport perspective, previous works that use an optimal transport map between two measure spaces in generative models [1, 24] have not focussed on the bilipschitz nature of WGANs making them perfect candidates for learning homeomorphisms. This serves as our motivation to study the effects of topological regularisation on the latent space of Generative Models like WGANs.

4 Proposed Method

Let $\mathcal{X} = \{x_0, x_1, \dots\}$ with $x_i \in \mathbb{R}^d$ denote the original data samples, which we consider to be samples from an underlying manifold $\mathbb{M}_{\mathcal{X}}$, and probability distribution $P(\mathcal{X})$. Let the generator of the proposed GAN be represented as $\mathcal{G}_\theta$ and the discriminator as $\mathcal{D}_\phi$. The major intuition that went inside this work is to disentangle the "disconnected support density estimation" problem into two parts, namely, (a) latent space topology optimization and (b) homeomorphic density estimation.

In the first part of the problem, we learn a mapping, parameterized by a Neural Network N_ψ, from $\mathcal{N}(0, 1)$ to a latent space with samples $\mathfrak{N} := \{n_0, n_1, n_2, \dots\}$, with $n_i \in \mathbb{R}^d$, corresponding to an underlying manifold $\mathbb{M}_\eta$ and distribution $P(\mathfrak{N})$. The objective of the optimization is to minimize the topological dissimilarity of $\mathbb{M}_\eta$ and $\mathbb{M}_{\mathcal{X}}$. More formally, we calculate the persistent homology (PH) of the Vietoris–Rips (VR) complex of $\mathcal{X}$, and $\mathfrak{N}$, denoted as $\mathfrak{R}_\epsilon(\mathcal{X})$ and $\mathfrak{R}_\epsilon(\mathfrak{N})$. The persistence diagrams $\mathcal{D}_{\mathcal{X}}$ and $\mathcal{D}_{\mathfrak{N}}$, obtained from PH calculation of the VR complex, encode the topological features of the underlying manifolds $\mathbb{M}_{\mathcal{X}}$ and $\mathbb{M}_{\mathfrak{N}}$, respectively. One straightforward way of calculating topological similarity between $\mathbb{M}_{\mathcal{X}}$ and $\mathbb{M}_\eta$, is to compare the persistence diagrams $\mathcal{D}_{\mathcal{X}}$ and $\mathcal{D}_{\mathfrak{N}}$ using some well known metric such as the 1-Wasserstein Metric, the Bottleneck-Distance, the Hausdorff Metric, etc. [8, 15, 18] However, these metrics do not provide gradients for samples in the sample space, precluding their use as an optimisation objective.A recent paper by Moor et al. [22], redefines the PH calculation to return multi-sets of persistence diagrams and persistence pairings, $\{\mathcal{D}_{\mathcal{X}}, \Pi_{\mathcal{X}}\}$ and $\{\mathcal{D}_{\mathfrak{N}}, \Pi_{\mathfrak{N}}\}$ respectively. Next it uses the Pairwise Distance matrix D between samples, and the persistence pairings $\Pi_{\mathcal{X}}$ and $\Pi_{\mathfrak{N}}$ to find out the topologically significant pairs of points and minimize their distance. This expression permits gradient calculations and is explicitly designed to be used as an objective function in topological regularization. The formal definition is shown by Eqn 2.
Given two sample spaces X and Y, with pairwise distance matrices, D_X and D_Y, the signature loss is defined as

$$\mathcal{L}_{signature}(X, Y) = \frac{1}{2} \left\| D_X[\Pi_X] - D_Y[\Pi_X] \right\|^2 + \frac{1}{2} \left\| D_X[\Pi_Y] - D_Y[\Pi_Y] \right\|^2 \tag{2}$$

The second and more complex part of the problem is to learn a homeomorphism between the distributions $P(\mathcal{X})$ and $P(\mathfrak{N})$ using a Generative Adversarial Network. Put briefly, the generator $\mathcal{G}_\theta$ needs to learn the original data distribution $P(\mathcal{X})$ while explicitly preserving the number of d-dimensional holes on $\mathbb{M}_\eta$ learned from the previous optimization routine.

According to [26], if a function f, defined on $\mathbb{R}^d \to \mathbb{R}^d$, is bi-lipschitz (i.e., f and f^{-1} are k-lipschitz continuous) or is an isometry (i.e., having lipschitz constant $L = 1$), it can be called a homeomorphism. With that said, various works have been done in the direction of imposing explicit isometry constraints on neural networks. Arjovsky et al. [3] achieved this by performing gradient clipping in discriminators, Gulrajani et al. [11] used a gradient penalty term to prevent strong gradients. Another work by Miyato et al. [21] imposes an explicit 1-Lipschitz constraint by spectral normalization of the weights of a network. This method has also found application in the Invertible Neural Network [5] literature, where it is an absolute necessity to have $L \leq 1$ Lipschitz continuity to maintain invertibility. In this work, we adapt the spectral normalization routine proposed by [21], to constrain $\mathcal{G}_\theta$ to be 1-Lipschitz continuous. More formally, for each training iteration, the weight matrices θ of the generator $\mathcal{G}_\theta$ is rescaled as $\theta = \frac{\theta}{\sigma_{\max}(\theta)}$ where $\sigma_{\max}(\theta)$ denotes the maximum singular value of θ. Along with the weight scaling, we use 1-lipschitz activations like, ReLU[12], Leaky ReLU[20], Tanh, Sigmoid, etc. for maintaining $L \leq 1$ throughout the layer compositions of the network architecture. This continuity constraint along with the equal dimensionality of the input and output space of $\mathcal{G}_\theta$, imposes the desired explicit homeomorphism constraint.

The final architecture (See Figure 2) consists of an encoder, N_ψ before the generator $\mathcal{G}_\theta$. The latent vectors sampled from N_ψ is sent into $\mathcal{G}_\theta$ in place of samples from $\mathcal{N}(0, 1)$. The Signature loss (Equation 2), is then used to compare the latent space topology and the topology of the desired data manifold. It has been observed, that min-max scaling the output of Noise Module between -1.0 and 1.0 aids optimization. We experiment on the Vanilla GAN architecture [10], and the WGAN-GP architecture [3, 11].

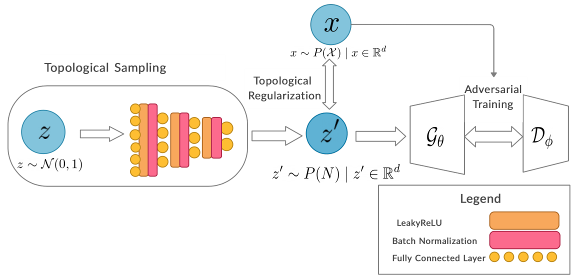


Figure 2: Entire Architecture with proposed topological sampling module

In both the cases, the discriminator loss (along with the gradient penalty term in case of WGAN-GP) remains the same. The generator loss for the vanilla GAN is re-defined as follows:

$$\mathbb{E}_{z' \sim P(\mathfrak{N})}\left[\ln(1 - \mathcal{D}_\phi(\mathcal{G}_\theta(z')))\right] + \left[\frac{1}{B}\sum \mathcal{L}_{signature}(N_\psi(z), x_r)\right]_{z \sim \mathcal{N}(0,1), x_r \sim P(\mathcal{X})} \tag{3}$$

Similarly for WGAN-GP we use optimize the following objective for the generator.

$$\mathbb{E}_{z' \sim P(\mathfrak{N})}\left[\mathcal{D}_\phi(\mathcal{G}_\theta(z'))\right] + \left[\frac{1}{B}\sum \mathcal{L}_{signature}(N_\psi(z), x_r)\right]_{z \sim \mathcal{N}(0,1), x_r \sim P(\mathcal{X})} \tag{4}$$

Here B denotes the mini-batch size being used, $P(\mathfrak{N})$ denotes the learned probability distribution of N_ψ. In both the scenarios, $z \sim \mathcal{N}(0,1)$ and $x_r \sim P(\mathcal{X})$, B samples are taken from the distributions $\mathcal{N}(0,1)$ and $P(\mathcal{X})$ respectively, sampled without replacement.

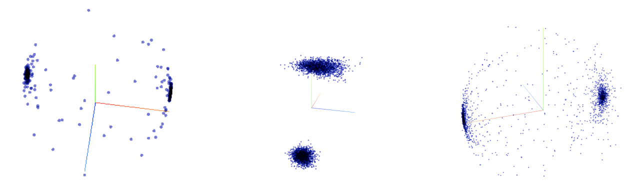
5 Experiments

In this section, we briefly discuss our preliminary experiments and their results. We are currently testing out MLP GANs on MNIST dataset, and using the Precision-Recall metrics for GANs [17] for assessing performance. We use the last layers of a pretrained Convolutional MNIST classifier. as the feature extractor. For the experiments, we are using a patched version of Ripser++ [27] for the PH calculations, to calculate Persistence Pairings along with the Persistence Diagrams.

In the experimental setup, we set the values of the hyperparameters as $\gamma = 10$, $\lambda = 0.01$, $lr_{\mathcal{G}_\theta} = 0.0002$, $lr_{\mathcal{D}_\phi} = 0.0004, B = 100$, epochs = 200. Our homeomorphism assumption of the generator has been split in two directions. A recent work by Li et al. [19] mentions "`GAN and VAE trainings approximate a homeomorphism between the data space and the sampling space`". However, Brehmer and Cranmer [6] discusses how GANs and VAEs consider the data to be embedded in a lower-dimensional manifold. This change in dimensionality violates the properties for the generators to be a homeomorphism. For this reason, we experiment on both these interpretation of "homeomorphism", namely low dimensional manifold embedding, and a properly defined homeomorphism from same-dimensional prior space to posterior space of $\mathcal{G}_\theta$.

To test the effectiveness of topological regularization, we propose a learning problem, with an objective to model a Gaussian mixture with disconnected supports. More formally, the target distribution $P(\mathcal{X})$ is defined as a mixture of Gaussians $\mathcal{N}(\mu_0, \sigma_0)$ and $\mathcal{N}(\mu_1, \sigma_1)$, $\|\mu_1 - \mu_0\|_1 \geq 3(\sigma_0 + \sigma_1)$ (by Pukelsheim [23]) such that there exists little-to-no samples which belonged to both these distributions. To prevent large variance in sample magnitudes, the generated samples are min-max scaled between $[-1, 1]$. In our experiments, we generated $60,000$ samples from the mixture, with $\mu_0 = 0$, $\sigma_0 = \sigma_1 = 1$ and $\mu_1 = 6$. To promote reuse of MNIST training setup with minimal changes in code, the mixture is set to generate i.i.d samples from $\mathbb{R}^{784}$. The final topological visualization is done by performing PCA[9] on the learned posterior of 784 dimensions.

Despite the intuitive theoretical background, the experimental results presented some surprising findings that challenged our initial assumptions. As shown in 3(c), the global topology of the learned prior obtained by the topology-constrained Noise Module, does not contain two distinctly separated



(a) (b) (c)

Figure 3: Figure 3(a) Posterior space topology of GANs when an isotropic Gaussian prior is used for learning a data distribution with disconnected supports. 3(b) Posterior space topology of GAN trained on the same dataset but with a disconnected prior. Note the absence of out of manifold samples. 3(c) Global topology of learned prior space.

Table 1: Experiments

MNIST			
GAN Type	Input Space Dim	Precision	Recall
Vanilla GAN	100	98.23%	81.2%
Vanilla GAN (with Topo Loss)	100	97.78%	80.54 %
Vanilla GAN (with Noise Layer)	100	90.2%	78.24%
Vanilla GAN (with Noise Layer)	784	78%	67.3%
WGAN (with Noise Layer and Spectral Generator)	784	**Mode Collapse**	
GAN (with Noise Layer and Spectral Generator)	784	**Mode Collapse**	

distributions (like the one in 3(b)). Instead we find sparsely-spread samples between two clusters. We suspect that this might be caused by the continuous nature of neural networks, due to which instead of introducing holes, $\mathbb{M}_\eta$ gets "stretched" between the two modes of $P(\mathcal{X})$, thus reducing out of manifold samples in between the modes, and increasing sample density near the modes.

The proposed GAN architecture also presented some counter-intuitive results. For starters, it has been noticed that introducing spectral normalization in $\mathcal{G}_\theta$ causes mode collapse in the GAN. Despite the poor quality of the samples, our initial experiments presents the importance of prior space topology for disconnected manifold learning. Figure 3(b) shows the GAN posterior space topology when trained with disconnected Gaussian mixture prior. The same model, when trained with an isotropic Gaussian prior (i.e. the original vanilla GAN formulation) produces out of-manifold samples, which are found sparsely spread between the learned clusters, as shown in figure 3(a).

Table 1 reports the comparison of precision-recall scores between our proposed architecture and regular GAN architectures like WGAN and vanilla JSD GAN [3, 10].

6 Conclusion

This work studies the problem of learning disconnected sample space manifolds in GANs by topologically aligning the prior space to the original data space. We introduce a persistent homology perspective towards augmenting the prior distribution to stay in the same homology class as that of our unknown data manifold. Although the empirical results do not yet support the ability of neural networks in introducing d-dimensional holes, we found that they provided some useful insights into the topological behaviour of neural networks. In a follow-up of this work we will rethink our assumptions and apply these newly-obtained insights for introducing d-dimensional holes to non-manifold priors.

Acknowledgments and Disclosure of Funding

We would like to extend our deepest gratitude, to Dr. Bastian Rieck, D-BSSE, ETH Zurich, for his unending support and mentorship throughout the process of experimentation and documentation. We would also like extend our gratitude to the reviewers for their valuable comments, helping us fix some important problems in the report. This project is self-funded by the authors, with no external funding / computation support from third-party sponsors.

References

[1] Dongsheng An, Yang Guo, Min Zhang, Xin Qi, Na Lei, Shing-Tung Yau, and Xianfeng Gu. Ae-ot-gan: Training gans from data specific latent distribution, 2020.

[2] Anonymous. Learning disconnected manifolds: Avoiding the no gan's land by latent rejection. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/pdf?id=nxJ8ugF24q2. under review.

[3] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein gan, 2017.

[4] S. Barannikov. The framed morse complex and its invariants. In *Singularities and Bifurcations*, 1994.

[5] Jens Behrmann, Will Grathwohl, Ricky T. Q. Chen, David Duvenaud, and Jörn-Henrik Jacobsen. Invertible residual networks, 2019.

[6] Johann Brehmer and Kyle Cranmer. Flows for simultaneous manifold learning and density estimation, 2020.

[7] H. Edelsbrunner and J. Harer. Persistent homology — a survey. In *Surveys on Discrete and Computational Geometry: Twenty Years Later*, 1994.

[8] A. Efrat, A. Itai, and M. J. Katz. Geometry helps in bottleneck matching and related problems. *Algorithmica*, 31(1):1–28, September 2001. doi: 10.1007/s00453-001-0016-8. URL https://doi.org/10.1007/s00453-001-0016-8.

[9] Karl Pearson F.R.S. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572, 1901. doi: 10.1080/14786440109462720.

[10] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks, 2014.

[11] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron Courville. Improved training of wasserstein gans, 2017.

[12] Richard H. R. Hahnloser, Rahul Sarpeshkar, Misha A. Mahowald, Rodney J. Douglas, and H. Sebastian Seung. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. *Nature*, 405(6789):947–951, Jun 2000. ISSN 1476-4687. doi: 10.1038/35016072. URL https://doi.org/10.1038/35016072.

[13] C. Hofer, F. Graf, M. Niethammer, and R. Kwitt. Topologically densified distributions. *arXiv*, 2020. https://arxiv.org/abs/2002.04805.

[14] Danijela Horak, Simiao Yu, and Gholamreza Salimi-Khorshidi. Topology distance: A topology-based approach for evaluating generative adversarial networks, 2020.

[15] Michael Kerber, Dmitriy Morozov, and Arnur Nigmetov. Geometry helps to compare persistence diagrams. *ACM J. Exp. Algorithmics*, 22, September 2017. ISSN 1084-6654. doi: 10.1145/3064175. URL https://doi.org/10.1145/3064175.

[16] Mahyar Khayatkhoei, Ahmed Elgammal, and Maneesh Singh. Disconnected manifold learning for generative adversarial networks, 2019.

[17] Tuomas Kynkäänniemi, Tero Karras, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Improved precision and recall metric for assessing generative models, 2019.

[18] Théo Lacombe, Marco Cuturi, and Steve Oudot. Large scale computation of means and clusters for persistence diagrams using optimal transport, 2018.

[19] Henry Li, Ofir Lindenbaum, Xiuyuan Cheng, and Alexander Cloninger. Variational diffusion autoencoders with random walk sampling, 2020.

[20] Andrew L. Maas, Awni Y. Hannun, and Andrew Y. Ng. Rectifier nonlinearities improve neural network acoustic models. In *in ICML Workshop on Deep Learning for Audio, Speech and Language Processing*, 2013.

[21] Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. Spectral normalization for generative adversarial networks. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=B1QRgziT-.

[22] Michael Moor, Max Horn, Bastian Rieck, and Karsten Borgwardt. Topological autoencoders, 2020.

[23] Friedrich Pukelsheim. The three sigma rule. *The American Statistician*, 48(2):88–91, 1994. doi: 10.1080/00031305.1994.10476030. URL https://amstat.tandfonline.com/doi/abs/10.1080/00031305.1994.10476030.

[24] Tim Salimans, Han Zhang, Alec Radford, and Dimitris Metaxas. Improving gans using optimal transport, 2018.

[25] Ugo Tanielian, Thibaut Issenhuth, Elvis Dohmatob, and Jeremie Mary. Learning disconnected manifolds: a no gans land, 2020.

[26] Jeremy T. Tyson. Handout on homeomorphisms, bi-lipschitz maps and isometries, 2006. URL https://faculty.math.illinois.edu/~tyson/homeo.pdf.

[27] Simon Zhang, Mengbai Xiao, and Hao Wang. Gpu-accelerated computation of vietoris-rips persistence barcodes, 2020.