filtrations have the advantage that they are usually efficient even for big point clouds, although they may be more challenging to use effectively. Given a point cloud (P, d), a subset of *landmarks* $L \subseteq P$, and $\alpha \geq 0$, the α-witness complex of (P, d) with landmarks L is the simplicial complex

$$W(P, L, \alpha, d) = \left\{\sigma \subseteq L : \exists w \in X \ \forall l \in \sigma \ \forall l' \in L \setminus \sigma, \ d(w, l)^2 \leq d(w, l')^2 + \alpha\right\}. \tag{11}$$

Choosing $\alpha_{\max} > 0$, we have that $W_{\alpha_{\max}} = (W(P, L, \alpha, d))_{\alpha \in [0, \alpha_{\max}]}$ is a filtration of simplicial complexes that induces a persistence module $\mathbb{V}_k(W_{\alpha_{\max}})$ where, for $t > \alpha_{\max}$, all the vector spaces forming the persistence module are equal to the one corresponding to the value $t = \alpha_{\max}$.

To measure the difference between the two persistence modules induced by the spaces $\mathcal{M}_X$ and $\phi_{\mathcal{G}}(\text{supp}(\mathbb{P}_Z))$, Khrulkov and Oseledets define the relative living times RLT of i-features for k-dimensional homology, given by

$$\text{RLT}(i, k, P, L, d) = \frac{\mu\left(\{\alpha \in [0, \alpha_{\max}] : \beta_k(W(P, L, \alpha, d)) = i\}\right)}{\alpha_{\max}}. \tag{12}$$

Note that the persistence modules, and thus the relative living times, are dependent on a choice of landmarks L. To minimize the effect of this dependency, Khrulkov and Oseledets set a fixed size of the set of landmarks L and compute the mean relative living times, given by the expectation of the relative living times over all possible choices of landmarks L of the fixed size, i.e., $\text{MRLT}(i, k, P, d) = \mathbb{E}_L[\text{RLT}(i, k, P, L, d)]$. Finally, the Geometry Score of a generative adversarial network $\mathcal{G}$ is given by the squared differences of all the mean relative living times for all the i for the 1-dimensional persistence modules generated by samples P and P' equipped with the same dissimilarity function d from $\mathcal{X}$ and $\phi_{\mathcal{G}}(\text{supp}(\mathbb{P}_Z))$, respectively, that is,

$$\sum_{i=0}^{\infty} \left(\text{MLRT}(i, 1, P, d) - \text{MLRT}(i, 1, P', d)\right)^2. \tag{13}$$

The validity of the Geometry Score as a quality metric was proved in several experiments. For example, it was ratified that the WGAN-GP (Gulrajani et al., 2017) is better than the WGAN model (Arjovsky et al., 2017), a well-known result, for the MNIST dataset. Additionally, the Geometry Score was able to distinguish between good and bad generative models derived from the DCGAN architecture (Radford et al., 2016) trained on the CelebA dataset (Liu et al., 2015).

In a very similar fashion, but with a simpler approach, Charlier et al. (2019) propose to compare generative models, this time not necessarily restricted to GAN models, by directly comparing the zero and one-dimensional persistence diagrams induced by Vietoris–Rips filtrations of samples taken from the real data distribution and from the generative model using the bottleneck distance. In this case, the experiments compared four different generative models in the credit card fraud detection dataset (cre). These models were WGAN and WGAN-GP together with WAE (Tolstikhin et al., 2018) and VAE (Kingma and Welling, 2014). In this case too, WGAN-GP was shown to produce better results than the other models.

The two aforementioned approaches evaluating the quality of generative models were based on comparing the topology of real and synthetic data by means of persistent homology. However, they are not the usual choices in the generative deep learning community, where scores such as the Fréchet inception distance (Heusel et al., 2017) or a numerical approximation of precision and recall (Sajjadi et al., 2018) are currently preferred to evaluate generative models. Kim et al. (2023) propose an alternative way to approximate precision and recall scores using an approximation of the support of the data distributions based on preimages of kernel density estimators (KDE) that preserve the topology of the support of a smoothed version of the real and synthetic data distributions. Their approximation is based on results of Fasy et al. (2014, Method IV), where it is argued that persistence diagrams from superlevel set persistence modules may carry topological information from the support of the data distribution. Specifically, the supports of the real and synthetic data are approximated by the superlevel sets $(\widehat{p}_{h_r})^{-1}[c_r, \infty)$ and $(\widehat{p}_{h_s})^{-1}[c_s, \infty)$, respectively, where $\widehat{p}_{h_r}$ and $\widehat{p}_{h_s}$ are KDEs for real and synthetic datasets $\mathcal{D}_r$ and $\mathcal{D}_s$ given by

$$\widehat{p}_{h_\bullet}(x) = \frac{1}{|\mathcal{D}_\bullet|} \sum_{p \in \mathcal{D}_\bullet} \frac{1}{h^d} K\left(\frac{x-p}{h}\right),$$

where $h > 0$ and K are the bandwidth and kernel of the KDEs, selected beforehand, and c_r and c_s are confidence bands for a given significance value α, obtained using bootstrap over the datasets $\mathcal{D}_r$ and $\mathcal{D}_s$ with the condition that

$$\liminf_{|\mathcal{D}_\bullet| \to \infty} \mathbb{P}\left(\|\widehat{p}_{h_\bullet} - p_{h_\bullet}\|_\infty < c_\bullet\right) \geq 1 - \alpha,$$

where p_{h_r} and p_{h_s} are smoothed versions of the real and synthetic data distributions, respectively (Fasy et al., 2014, Equation (26)). By the stability theorems (Cohen-Steiner et al., 2007), the confidence bands c_r and c_s, also bound the distances between the persistence diagrams coming from the functions $\widehat{p}_{h_\bullet}$ and $p_{h_\bullet}$ with probability $1-\alpha$, allowing one to study which points in the persistence diagrams generated from the data samples $\mathcal{D}_\bullet$ by $\widehat{p}_{h_\bullet}$ can be considered noise with respect to the persistence diagrams generated by $p_{h_\bullet}$; see Fasy et al. (2014, Section 4). The superlevel sets $(\widehat{p}_{h_\bullet})^{-1}[c_\bullet, \infty)$ are regions which intend to induce persistence diagrams without noisy points according to the previous confidence bands, thus recovering supports with similar topology to the topology of the support of the smoothed data distributions measured by the distances between their respective persistence diagrams.

Under some technical assumptions, the precision and recall scores approximated using the superlevel sets proposed by Kim et al. become close to the real precision and recall from the distribution as more examples are added to the datasets $\mathcal{D}_r$ and $\mathcal{D}_s$. The approximation is done with robustness, meaning that it holds even with data possibly corrupted by noise. The suitability of the new metrics was tested on several synthetic and real scenarios, where the metric accurately describes the differences between real and synthetic data distributions in several scenarios where other metrics struggle to evaluate differences. Also, the F1-score computed from the precision and recall approximations ranked different generative models such as StyleGAN2 (Karras et al., 2020), ReACGAN (Kang et al., 2021), among others, as the FID metric, the primary metric to score generative models in the moment of the publication of this survey, making this metric consistent with the state-of-the-art knowledge about the quality of generative models.

A class of generative models create synthetic data transforming values from a *latent space* $\mathcal{Z}$ to the real data space $\mathcal{X}$ via a function $G\colon \mathcal{Z} \to \mathcal{X}$. For such generative models, a desirable property is to be able to control the attributes of the generated samples. One approach to controlling attributes involves designing generative models that map from the latent space to the data space in such a way that the latent space $\mathcal{Z}$ can be factored into subspaces corresponding to *factors of variation* of the data generated, meaning that if one changes the value of a subspace associated with a factor of variation in the latent space, then the generated data change only in this source of variation. A model that satisfies this property is said to be *disentangled*. For a formal discussion of disentanglement, we refer the reader to Higgins et al. (2018). Preliminary work on disentangling generative models using topological data analysis to build regularization terms can be found in Balabin et al. (2023). In this work, regularization terms are based on the differences between generated data from two samples from the latent space using the representation topology divergence, a method to compare point clouds topologically which is reviewed in Section 3.3.2.

Measuring a model's entanglement is challenging, and there is no canonical way to do it. Zhou et al. (2021) propose two measures of disentanglement, one unsupervised and one supervised, using topological data analysis. Although they perform similarly to other disentanglement metrics, these measures do not require some strong assumptions about the dataset or the model, except for the (weak) hypothesis that $\mathrm{Im}(G)$ is a manifold for the unsupervised case and the more unrealistic assumption that the data space is a manifold for the supervised case. Roughly, these measures are based on two fundamental ideas for disentangled models: 1. The family of submanifolds of $\mathrm{Im}(G)$ yielded by the images of G restricted to the different values of the same subspace corresponding to a factor of variation are pairwise homeomorphic and thus have the same topology; 2. Submanifolds corresponding to different factors of variation are usually not homeomorphic.

The measures, based on the above ideas, penalize inconsistent topologies from the same factor of variation but reward topological differences between separate factors. However, the association of the latent dimensions of $\mathcal{Z}$ with the factors of variation is often unknown. Also, in practice, it is impossible to compute the full submanifolds for each restriction. To address this, the topological differences between the submanifolds of the different factors of variation are approximated by the topological differences of the submanifolds induced by the latent variables. Specifically, submanifolds are computed fixing values for the latent variables instead of for the factors of variation, and the topology for each submanifold is characterized by relative living times, that only need samples from the submanifolds. Then, Wasserstein barycenters (Agueh and Carlier, 2011) are calculated for each latent dimension from the relative leaving times, leading to a dissimilarity matrix M based on the pairwise Wasserstein distances between the Wasserstein barycenters. This matrix, processed by a (co)clustering algorithm, produces another dissimilarity matrix M' measuring the *topological similarities* of c different probable factors of variation, from which the unsupervised disentanglement measure is derived as

$$\mu = \mathrm{tr}(M') - \left(\sum_{i=1}^{c} \sum_{j=1}^{c} M'_{i,j} - \mathrm{tr}(M') \right). \tag{14}$$

The supervised version, assuming that the data space is already factored into subspaces corresponding to the factors of variation, differs by generating the matrix M, this time