# PHom-GeM: Persistent Homology for Generative Models

Jérémy Charlier

Last Year PhD Student at University of Luxembourg
Visiting PhD Student at Columbia University


# Outline


1. Introduction
   - Context
   - Research Question
2. Methodology
   - Persistent Homology Concepts
   - Persistent Homology for Generative Models
3. Experiments
   - Data Availability
   - Results
4. Conclusion

# Context

Generative models (GANs, AE) famous to generate adversarial samples

Samples quality measured by images generation

Figure 1: Visual sampling is a popular technique to measure the quality of artificially generated adversarial samples.

# Context

What can we do for non image-based applications?

Traditional distance measures fail to reflect intuitively the samples quality

Persistent homology specifically designed to describe data points cloud

# Research Question

How can we apply persistent homology to generative models
to assess the quality of adversarial samples
in real-world and non image-based applications?

Solution and Contributions

- A **P**ersistent **Hom**ology procedure for **Ge**nerative **M**odels
- The bottleneck distance measure for persistence diagrams
- Real-world application on credit card transactions

# Persistent Homology Concepts

Persistent Homology

- describes the shape of the data points cloud
- relies on features such as connected components, loops or cavities
- is independent of any distance measurement

Categorization into different homology groups

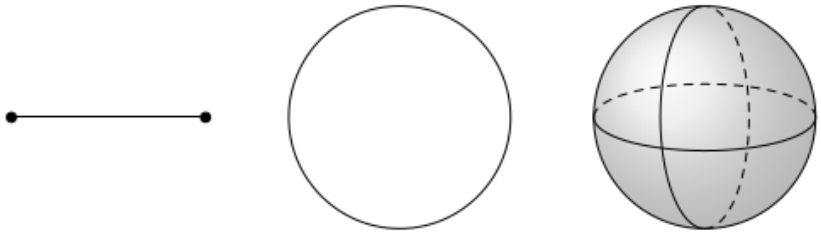


Figure 2: Visualization of the first three homology groups $H_0$, $H_1$ and $H_2$.

# Persistent Homology Concepts

Simplicial complex

- is a collection of numerous "simplex"
- is used to describe the homological properties of the data
- 0-simplex = point
- 1-simplex = line
- 2-simplex = triangle
- 3-simplex = tetrahedron

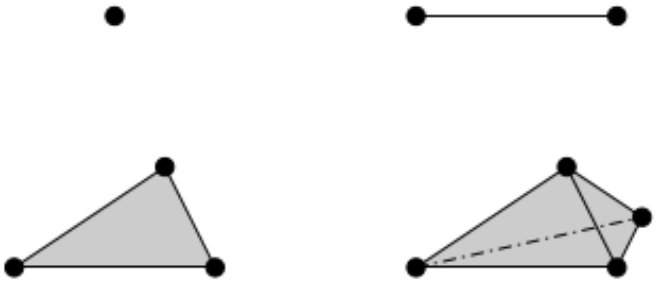


Figure 3: Visualization of different simplex.

# Persistent Homology Concepts

Filtration parameter $\varepsilon$

- $\varepsilon$ grows around each data point
- A line is drawn when two disks intersect
  - $\hookrightarrow$ Creation of 1-simplex
- Triangles are generated as $\varepsilon$ keeps growing
  - $\hookrightarrow$ Creation of 2-simplex

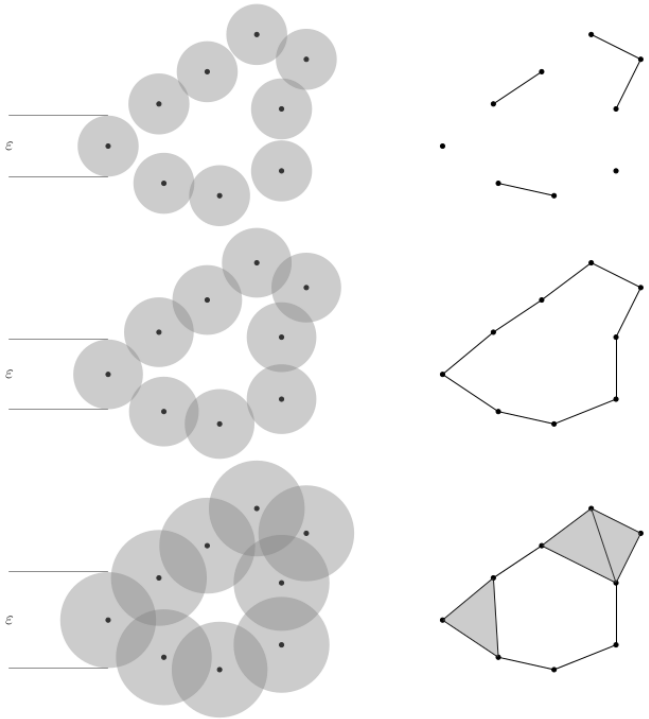



Figure 4: Filtration parameter growth and simplex construction.

# Persistent Homology Concepts

Barcodes and Persistence Diagrams

- highlight the persistent homology features
- describe the birth-death cycle
- Use of the bottleneck distance with the persistence diagrams
  - Characterize similarities between different diagrams

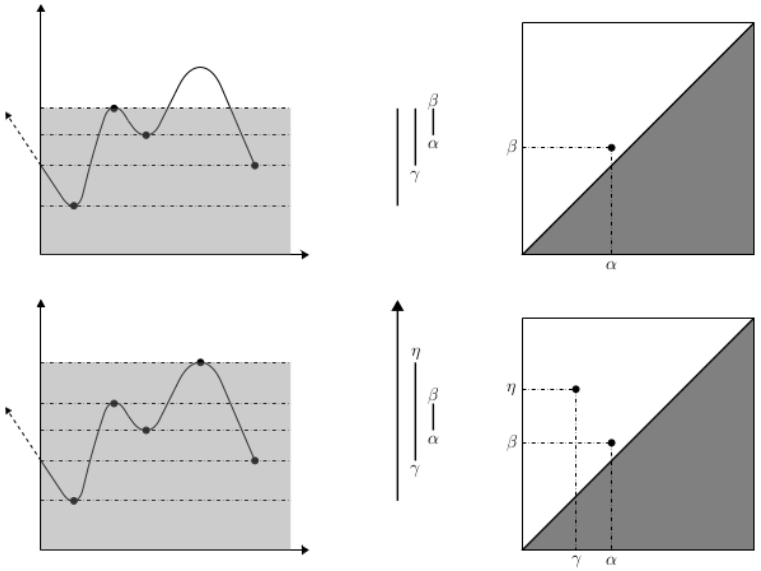



Figure 5: The local minima of the function provoke the creation of a barcode. The local maxima lead to the death of the barcode.

# Persistent Homology Concepts

Combining Filtration Parameter, Homology Groups and Barcodes

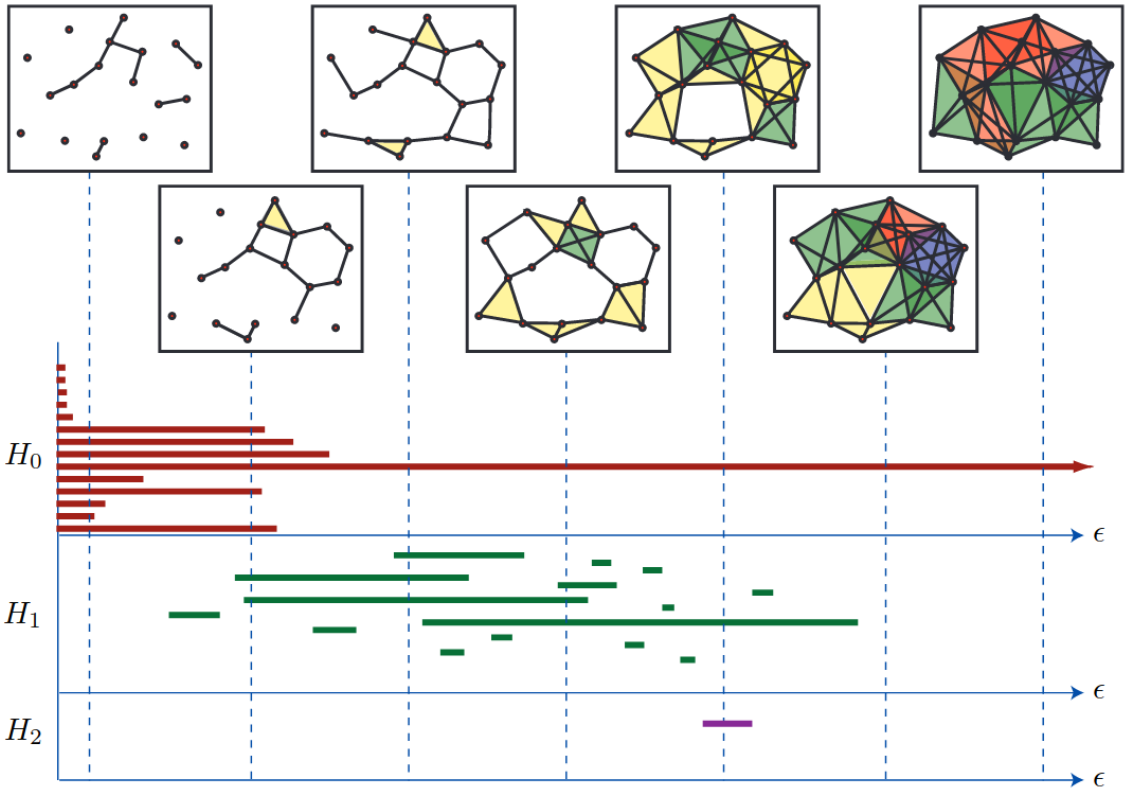



Figure 6: Persistent homology features for data points inherited from an annulus.

Persistent Homology for Generative Models applied to GANs

- Mapping of original and generated manifolds to metric space sets
- Creation of filtered simplicial complex
- Description of persistent homological features

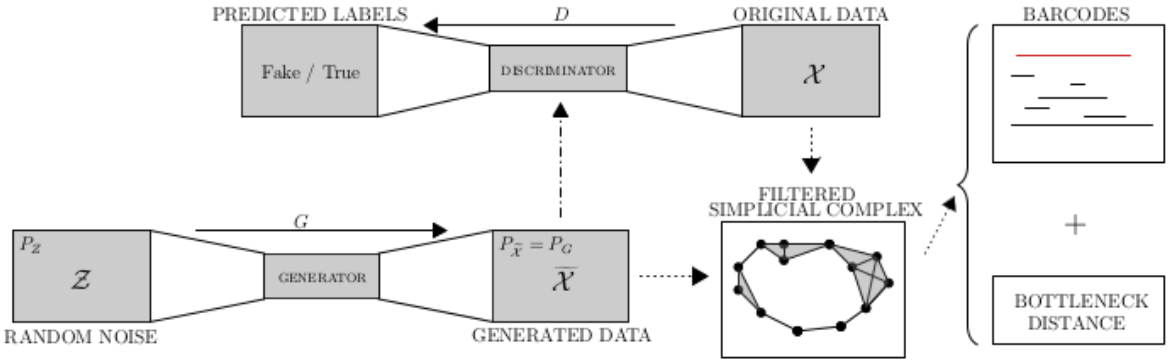



Figure 7: PHom-GeM applied to GANs.

# PHom-GeM

Persistent Homology for Generative Models applied to AEs

- Assess the persistent homological similarities between
  - the original and decoded data
  - the adversarial samples generated by the AE

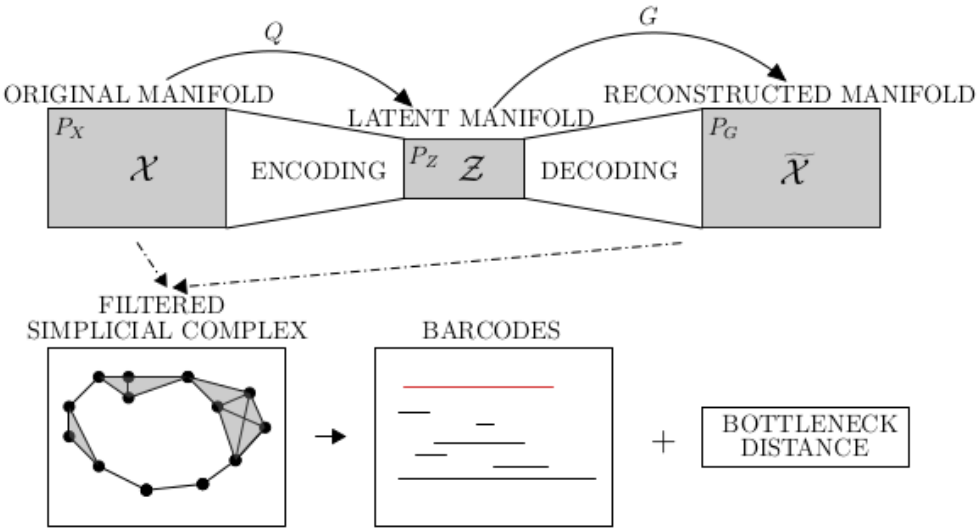



Figure 8: PHom-GeM applied to AEs.

# Data Availability

Use of a public data set

- Credit card transactions data set of the ULB Machine Learning Group
- Extracted from the Kaggle database
- `https://www.kaggle.com/mlg-ulb/creditcardfraud`

Overview of the data

- Anonymized data set
- 2 days of credit card transactions
- 29 features including the amount

# Results

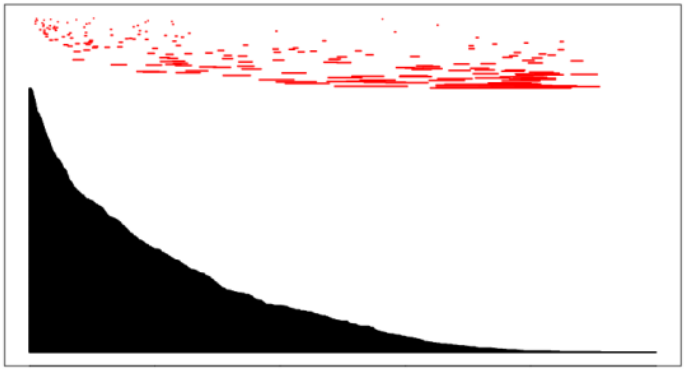


Figure 9: Original Sample

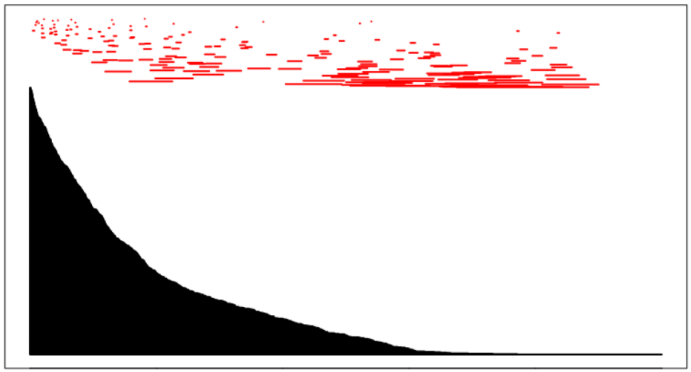


Figure 10: GP-WGAN

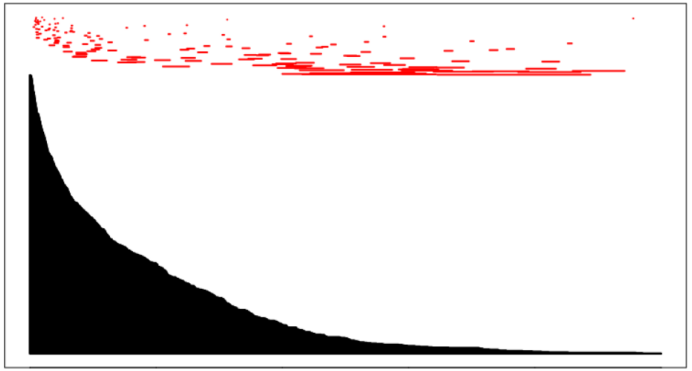


Figure 11: WGAN

# Results

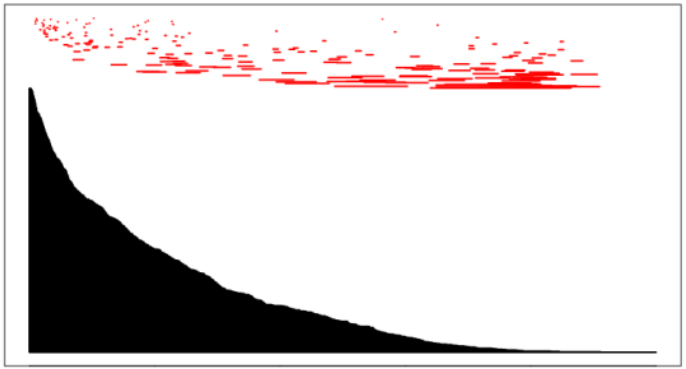


Figure 12: Original Sample

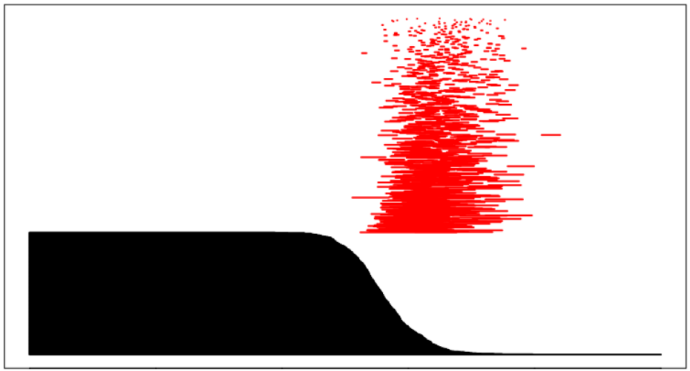


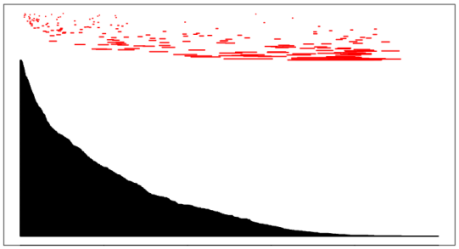
Figure 13: WAE

Figure 14: VAE

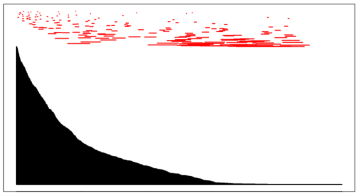
# Results

Comments

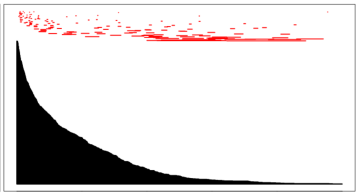
- Significant differences between GANs and AEs
- GANs better replicate the persistent homological features
- Spectrum of AEs barcodes is narrower



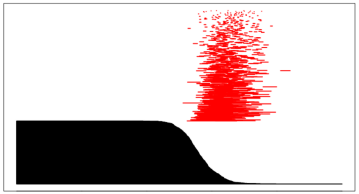
Original Sample



GP-WGAN



WGAN



WAE

VAE

Bottleneck distance for quantitative comparison

- Compare persistent homological similarities between the models
- Confirms the visual observations
- The lower, the better

| Gen. Model | Mean Value | Lower Bound | Upper Bound |
|---|---|---|---|
| GP-WGAN | **0.0711** | **0.0683** | **0.0738** |
| WGAN | 0.0744 | 0.0716 | 0.0772 |
| WAE | 0.0821 | 0.0791 | 0.0852 |
| VAE | 0.0857 | 0.0833 | 0.0881 |

Figure 18: Bottleneck distance between generated and original manifolds.

# Conclusion

Summary

- Persistent Homology for Generative Models
- Highlight the manifold features of the generative models for non image-based applications
- Experiments performed on a challenging credit card transactions data set
- In our configuration, GANs better preserve the persistent homological features
  - Qualitatively and quantitatively

# Conclusion

Future Work

- Influence of the homotopy type in the results
- Integrate a topological optimization function as a regularizer term

## Thank you for your attention

Jérémy Charlier

jeremy.charlier@uni.lu

www.linkedin.com/in/jeremy-charlier