Topological Data Analysis for Gravitational Wave Detection under Low Signal-to-Noise Ratios

Gabriel Wendell Celestino Rocha[1,*], Gandhimohan Madras Viswanathan[1]

[1]Dept. of Physics, Universidade Federal do Rio Grande do Norte,

1. Introduction

Gravitational-wave (GW) signals are often buried in detector noise, especially at low signal-to-noise ratios (SNR). Traditional pipelines rely on matched filtering and time–frequency representations, which exploit prior knowledge of the waveform family.

Here we explore *Topological Data Analysis* (TDA) as a complementary approach. Instead of focusing solely on amplitude or spectral content, we study the **shape** of GW-like time series in reconstructed phase space via delay embeddings and persistent homology.

2. Motivation

Key idea: GW-like chirps induce oscillatory dynamics. Under Takens delay embedding, such signals trace out loop- or torus-like manifolds in $\mathbb{R}^m$, while noise fills space more uniformly.

- Signals $\to$ coherent trajectories $\to$ non-trivial H_1 (loops).
- Noise $\to$ amorphous clouds $\to$ short-lived features near the diagonal.

TDA offers **stable, geometric and topological invariants** (Betti numbers, persistence diagrams) that distinguish these regimes.

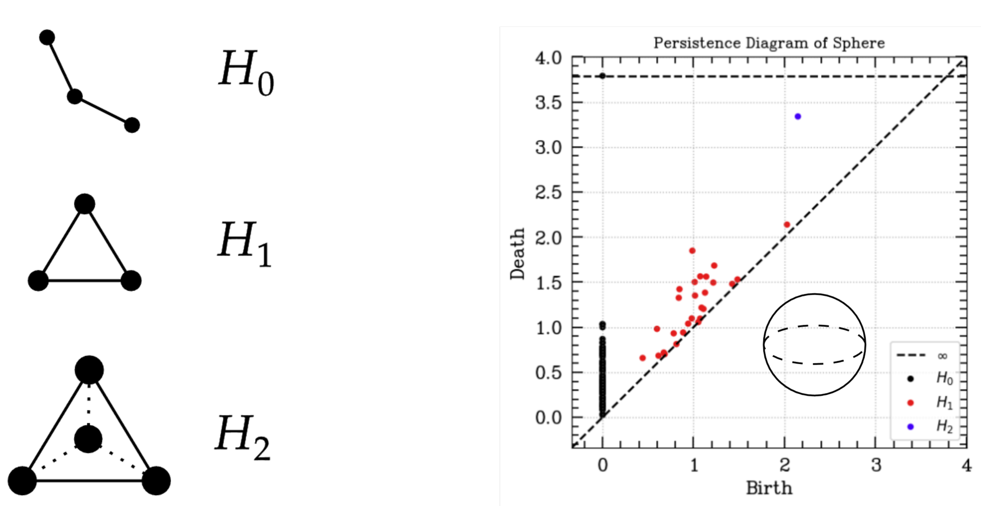
3. Objectives

- Build a TDA pipeline for GW-like signal detection using synthetic data.
- Characterize how persistent homology (especially H_1) evolves with SNR and noise type.
- Compare topological feature families (PI, PL, BC) to classical baselines.
- Assess robustness to $\pm 10\%$ perturbations in (m, τ).
- Profile computational cost and identify bottlenecks.

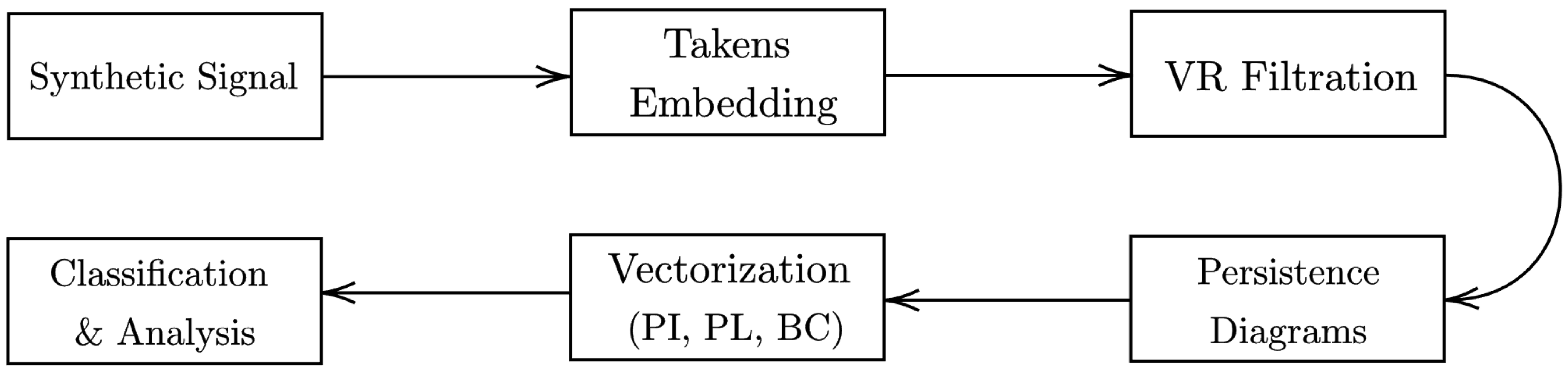
Takeaway: Detect GWs not only by "how loud" they are, but by *how they shape phase space*.

4. Mathematical Framework

$$\begin{cases} \Phi(t) = [x(t), x(t-\tau), \dots, x(t-(m-1)\tau)] \\ \mathcal{R}_\varepsilon(X) = \{[v_0, \dots, v_k] : d(v_i, v_j) \le \varepsilon\} \\ \mathcal{D}_k = \{(b_i, d_i)\}_i \quad , \quad p_i = d_i - b_i \quad , \quad k = 0, 1 \end{cases}$$

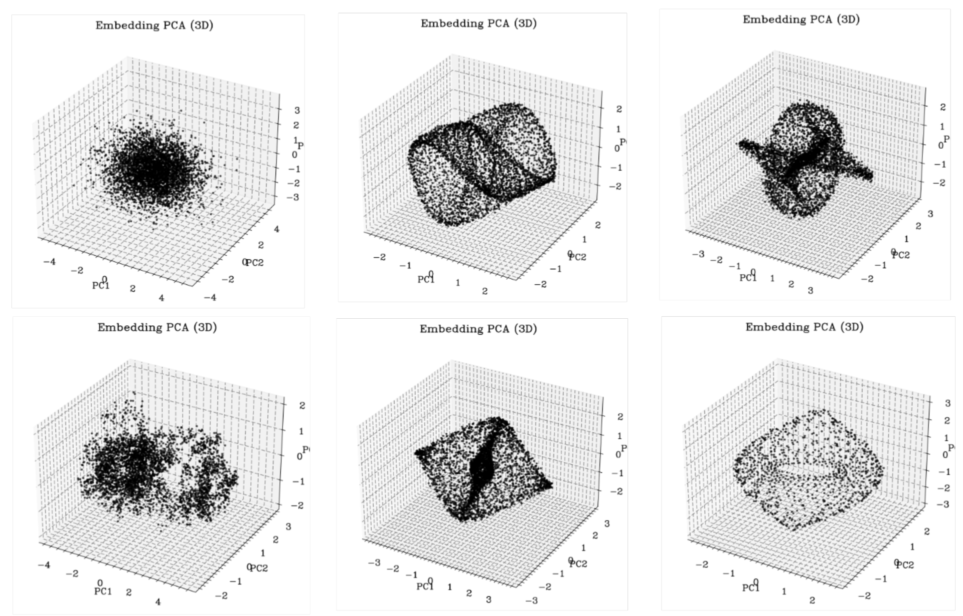


5. Pipeline Overview

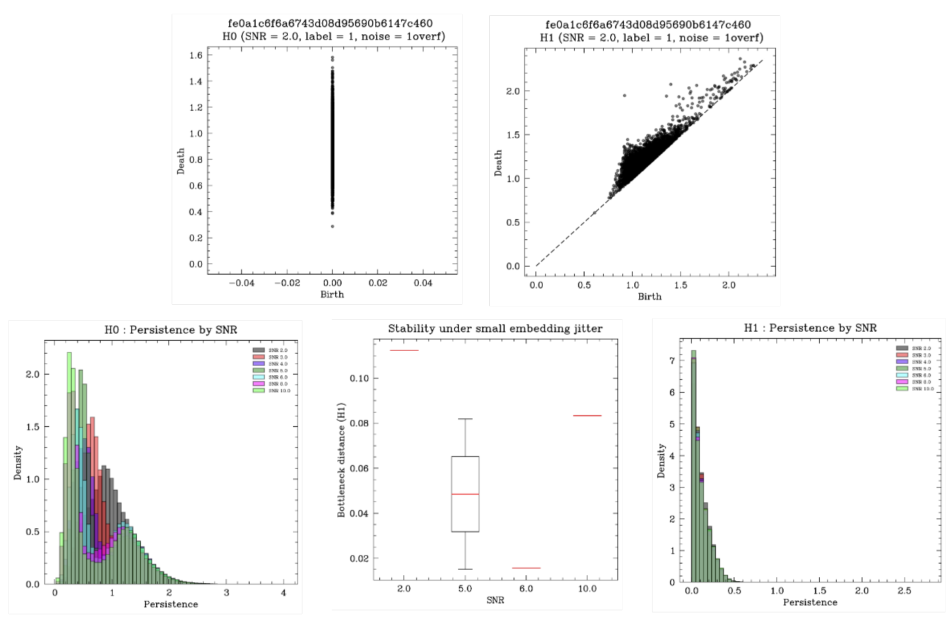


8. Results & Analysis

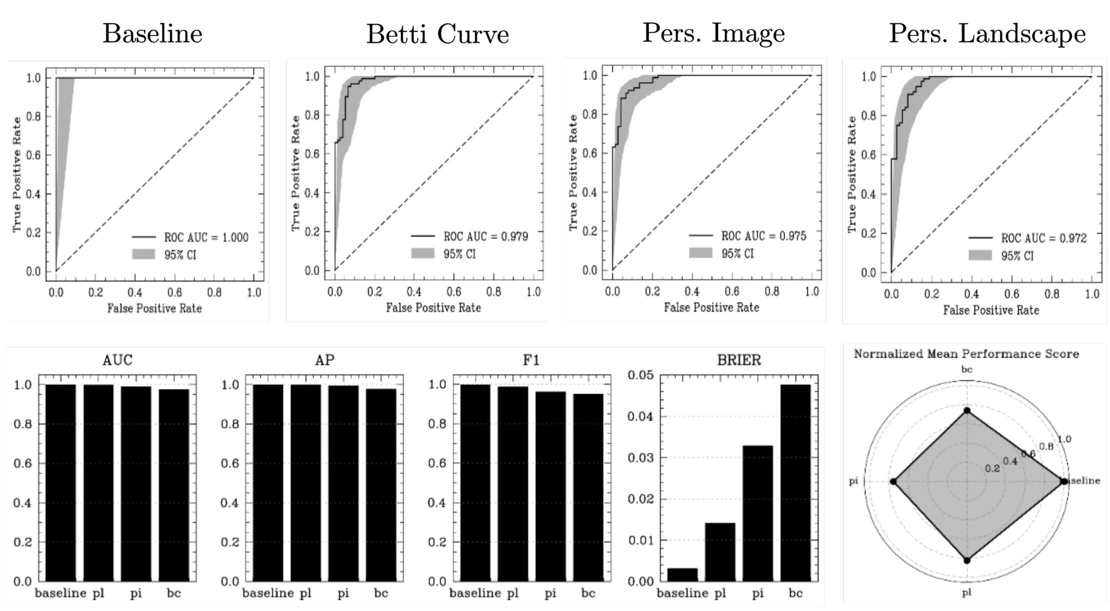
- **Embeddings**



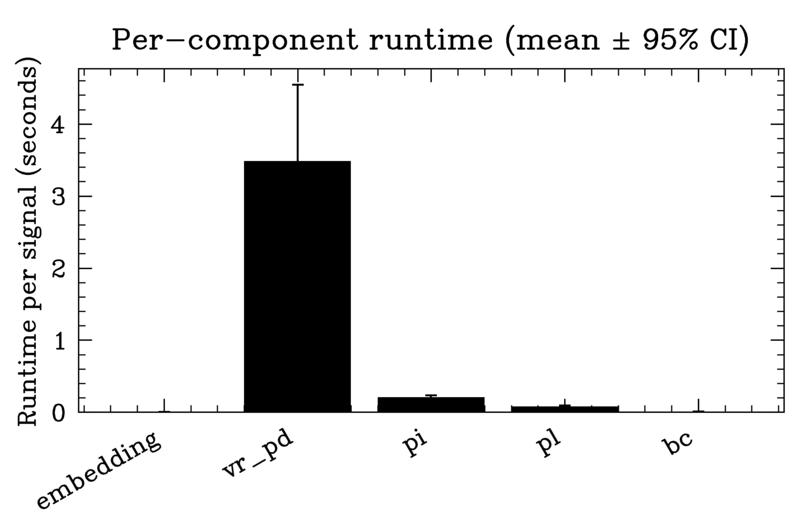
- **Diagnostic Analysis**



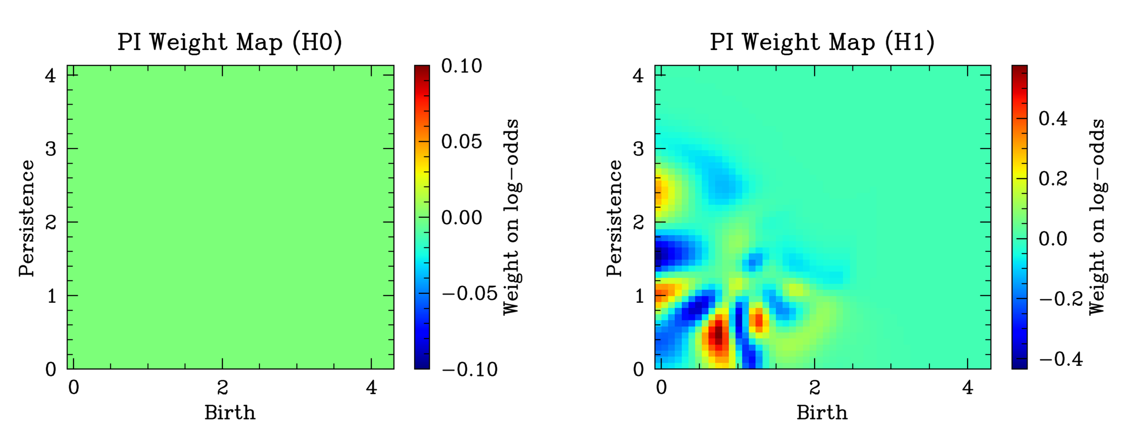
- **Performance Plots**



- **Runtime Breakdown**



- **PI Weight Maps**

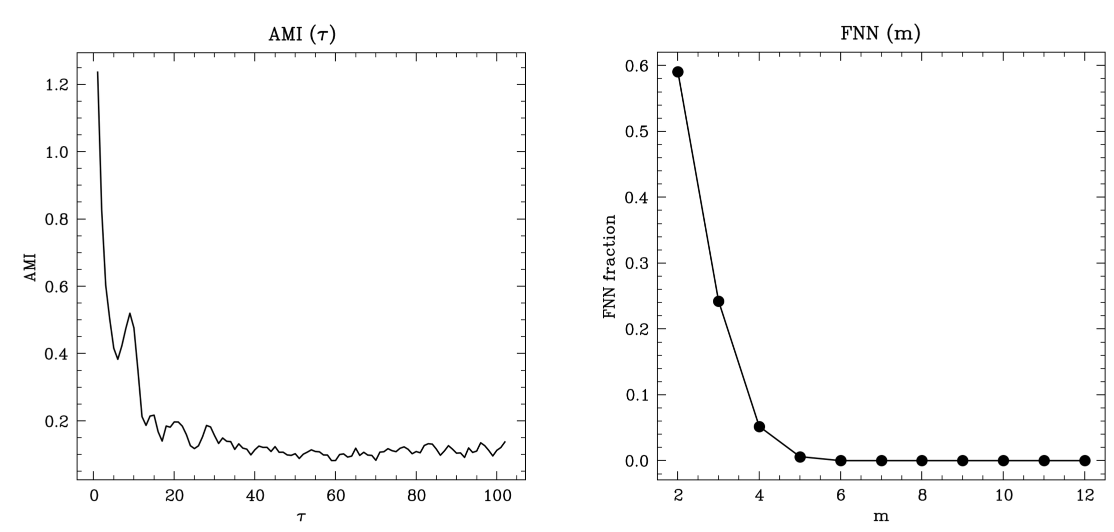


6. Embedding Parameters

1. Delay τ is chosen from the first local minimum of the **Average Mutual Information**:

$$\mathrm{AMI}(\tau) = \sum_{x,y} p_{x,y}(\tau) \log \frac{p_{x,y}(\tau)}{p_x p_y} .$$

2. Embedding dimension m is selected using **False Nearest Neighbors (FNN)**, i.e. the smallest m such that the fraction of false neighbors drops below a threshold.



7. Topological Feature Map

We transform persistence diagrams into fixed-length feature vectors:

$$I(x, y) = \sum_i w_i \exp\left[-\frac{(x - b_i)^2 + (y - p_i)^2}{2\sigma^2} \right] .$$

We train logistic regression (and other baselines) on:

- PI, PL, BC feature vectors.
- Classical baseline features.

The metrics used to evaluation were:

$$\mathrm{AUC}\ , \quad \mathrm{AP}\ , \quad F_1\ , \quad \mathrm{Brier} = \frac{1}{N}\sum_i (p_i - y_i)^2 .$$

- **Training:** high-SNR.
- **Testing:** includes low-SNR.

9. QR Code

For more details, please refer to this project's repository on GitHub using the QR code on the side.

- *gabrielwendell@fisica.ufrn.br