Antman Persistence: Detecting Small Holes with the Robust Density-Aware Distance (RDAD) Filtration

Chunyin Siu
cs2323@cornell.edu

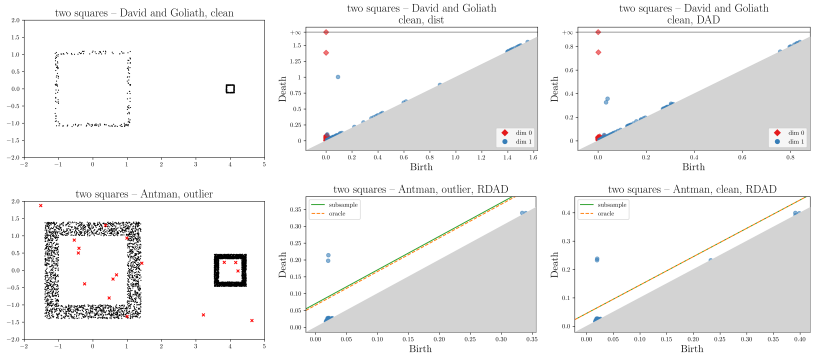
joint work with Gennady Samorodnitsky, Christina Yu and Andrey Yao
Center for Applied Mathematics, Cornell University

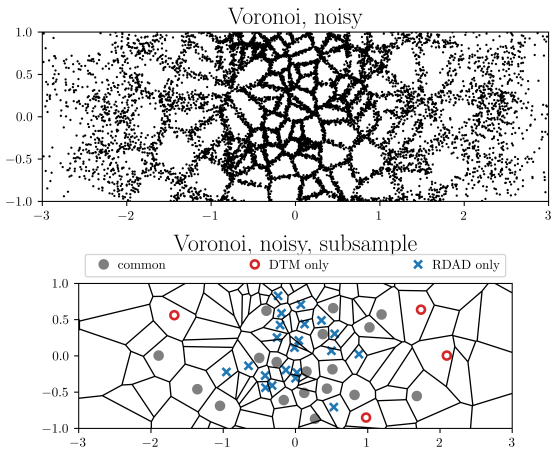
Proposed Filtration: RDAD (and DAD)

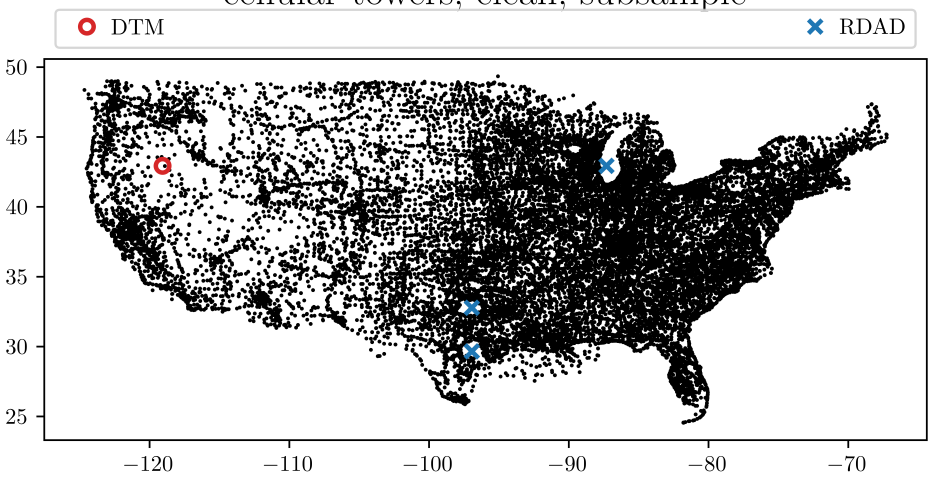
The traditional distance filtration can
- ✓ pick up big holes,
- ✗ but it cannot detect small holes.

The proposed filtration RDAD and DAD can
- ✓ pick up small holes.

In particular, RDAD can
- ✓ pick them up robustly and
- ✓ pick them up with a confidence band







What's the Math Behind?

Main ideas

- **DAD** scales distance with density, to magnify small holes surrounded by high-density regions (Bell et al 19, Hickok 22)
- **RDAD** considers averages of a few nearest points (by density-aware distance) (Chazal et al 18)
- **Bootstrapping** by subsampling from the dataset gives confidence bands (Chazal et al 18)

Nice properties

- Scale-invariance
- Robustness against Additive Noise and Outliers

What's the Definition?

- X: $\mathbb{R}^D$ random variable with density f
- DAD(x): infimum of the random variable $f(X)^{1/D}d(X,x)$.
- RDAD(x): average of the lowest quantiles of $f(X)^{1/D}d(X,x)$

$$RDAD(x;f,m) = \sqrt{\frac{1}{m}\int_0^m F_x^{-1}(q)^2 dq},$$

where $F_x(r) = P\left[f(X)^{1/D}d(X,x) \le r\right]$ and $m \in (0,1)$.

How to Estimate?

- use nearest neighbor distances to approximate the density
- use the average of the Nm smallest $f(X_i)^{1/D}d(X_i,x)$'s

Our paper
- C. Siu, G. Samorodnitsky, C. Yu, and A. Yao: Detection of small holes by the scale-invariant Robust Density-Aware Distance (RDAD) filtration (2022).
- Our codes: https://github.com/c-siu/RDAD

See our paper for the full bibliography.

Cited Works on this Poster
- Bell, G., Lawson, A., Martin, J., Rudzinski, J., Smyth, C.: Weighted persistent homology. Involve 12(5), 823–837 (2019).
- F. Chazal, B. Fasy, F. Lecci, B. Michel, A. Rinaldo, L. Wasserman: Robust topological inference: Distance to a measure and kernel distance. Journal of Machine Learning Research 18, 1–40 (2018).
- A. Hickok.: A Family of Density-Scaled Filtered Complexes (2022).