

## Abstract

In topology inference from data, current approaches face two major problems. One concerns the selection of a correct parameter to build an appropriate complex on top of the data points; the other involves with the typical ‘large’ size of this complex. We address these two issues in the context of inferring homology from sample points of a smooth manifold of known dimension sitting in an Euclidean space  $\mathbb{R}^k$ . We show that, for a sample size of  $n$  points, we can identify a set of  $O(n^2)$  points (as opposed to  $O(n^{\lceil \frac{k}{2} \rceil})$  Voronoi vertices) approximating a subset of the medial axis that suffices to compute a distance sandwiched between the well known *local feature size* and the local *weak feature size* (in fact, the approximating set can be further reduced in size to  $O(n)$ ). This distance, called the *lean feature size*, helps pruning the input set at least to the level of local feature size while making the data locally uniform. The local uniformity in turn helps in building a complex for homology inference on top of the sparsified data without requiring any user-supplied distance threshold. Unlike most topology inference results, ours does not require that the input is dense relative to a *global* feature such as *reach* or *weak feature size*; instead it can be adaptive with respect to the local feature size. We present some empirical evidence in support of our theoretical claims.