

Benjamin Eltzner

Torus-PCA with Application to Clustering of RNA Structures

Joint with Stephan Huckemann, Kanti V. Mardia

We propose a novel torus-PCA method that can be generally applied. Existing methods fall into two classes: tangent space PCA and geodesic PCA. However, unlike tangent space PCA, our torus-PCA honors the cyclic topology of the data space whereas, unlike geodesic PCA, our torus-PCA produces a variety of non-winding, non-dense descriptors. This is achieved by deforming tori into spheres and then using a variant of the recently developed principle nested spheres (PNS) analysis. We introduce a data-adaptive pre-clustering technique to optimize the deformation and a post-mode hunting technique for more fine-grained clustering if data are well approximated by a circle. Using our method to cluster a large RNA data set, we achieve sensible clusters containing more data points than previously used clustering schemes have found.