

Varun Khurana

University of California, San Diego, United States

vkhurana@ucsd.edu

We introduce LOT Wassmap, a computationally feasible algorithm to uncover low-dimensional structures in the Wasserstein space. The algorithm is motivated by the observation that many datasets are naturally interpreted as probability measures rather than points in \mathbb{R}^n , and that finding low-dimensional descriptions of such datasets requires manifold learning algorithms in the Wasserstein space. Most available algorithms are based on computing the pairwise Wasserstein distance matrix, which can be computationally challenging for large datasets in high dimensions. Our algorithm leverages approximation schemes such as Sinkhorn distances and linearized optimal transport to speed-up computations, and in particular, avoids computing a pairwise distance matrix. We provide guarantees on the embedding quality under such approximations, including when explicit descriptions of the probability measures are not available and one must deal with finite samples instead. These approximation guarantees provide a framework for using different finite sample approximations of the optimal transport map even when the underlying distributions are supported on all of \mathbb{R}^n . Experiments demonstrate that LOT Wassmap attains correct embeddings and that the quality improves with increased sample size. We also show how LOT Wassmap significantly reduces the computational cost when compared to algorithms that depend on pairwise distance computations.

Joint work with Keaton Hamm (University of Texas at Arlington, Arlington, TX), Alex Cloninger (University of California, San Diego, CA) and Caroline Moosmueller (University of North Carolina at Chapel Hill, NC).