Multiscale Geometric Methods for Statistical Learning and Data in High-Dimensiona

Mauro Maggioni, Duke University

We discuss a family of ideas, algorithms, and results for analyzing various new and classical problems in the analysis of high-dimensional data sets. These methods rely on the idea of performing suitable multiscale geometric decompositions of the data, and exploiting such a decomposition to perform a variety of tasks in signal processing and statistical learning. In particular, we discuss the problem of dictionary learning, where one is interested in constructing, given a training set of signals, a set of vectors (dictionary) such that the signals admit a sparse representation in terms of the dictionary vectors. We discuss a multiscale geometric construction of such dictionaries, its computational cost and online versions, and finite sample guarantees on its quality. We then generalize part of this construction to other tasks, such as learning a an estimator for the probability measure generating the data, again with fast algorithms with finite sample guarantees, and for learning certain types of stochastic dynamical system in high-dimensions.