

Invited Talk Abstract

ICOML 2026 | July 27–29, 2026

Point Sketches versus Coresets in k -Means**Leo Liberti***LIX, CNRS and Ecole Polytechnique, Institut Polytechnique de Paris*

Date	July 27, 2026
Time	16:30–17:00
Session	Session 3
Venue	Department of Mathematics, National Taiwan Normal University, Taipei, Taiwan

Abstract

Given a set of n points in d dimensions, the k -means problem consists in finding a clustering of given cardinality k such that the sum of squared Euclidean distances from each given point to the centroid of the cluster it belongs to is minimum. The problem inherits its name from a homonymous heuristic algorithm that is known to be both fast and effective. Random projections have been applied to the k -means problem to decrease the dimensions of each point, while coresets have been used to decrease the number of points. We show that the latter feat can also be achieved through a matrix sketch—a construction similar to random projections—applied to the rows of the data matrix, i.e. the given points in a k -means instance. Our computational tests indicate that, on single values of k , coresets give better quality results faster: but point sketches become much faster than coresets when one needs to rapidly compute approximate clusterings for many values of k .