

# A New Algorithm for Robust Affine-Invariant Clustering

Michael Pokojovy<sup>\*,†,‡</sup> and Andrews T. Anum<sup>†</sup>

<sup>\*</sup>Department of Mathematical Sciences

<sup>†</sup>Data Science Program

<sup>‡</sup>Computational Science Program

The University of Texas at El Paso, El Paso, TX

## Abstract

Cluster analysis is an unsupervised machine learning technique commonly employed to partition a dataset into distinct categories referred to as clusters. The  $k$ -means algorithm is a prominent distance-based clustering method. Despite overwhelming popularity, the algorithm is neither invariant under non-singular linear transformations nor robust, i.e., can be unduly influenced by outliers. To address these deficiencies, we adopt an alternative clustering procedure based on minimizing a “trimmed” variant of the negative log-likelihood function and develop a novel “concentration step” (C-step) that can iteratively reduce the objective function. A simulation study over multiple synthetic scenarios and a real-world example are analyzed to assess the performance of our algorithm. Compared to  $k$ -means, empirical studies indicate competitiveness and oftentimes superiority of our algorithm.

**Keywords:** Cluster analysis; affine invariant estimator; robust estimator