Heuristic aspects of exact algorithms for minimum sum-of-squares clustering

Daniel Aloise – École Polytechnique de Montréal
Pierre Hansen – GERAD and HEC Montréal

The problem of partitioning into M classes, or clusters, a given set of N points in the Euclidean space in order that the sum of squared distances from each point to the centroid of its cluster is minimized, is central to Cluster Analysis. Indeed, several thousand papers on this problem, its variants and applications have appeared during the last fifty years. Most algorithms proposed are heuristics, often following in recent years the framework of some metaheuristic. Exact methods are based on (i) branch-and-bound, (ii) linear programming and column generation, (iii) reformulation-linearization technique and (iv) spectral bounds. Each time heuristics choices have to be made and strongly influence the exact algorithm's performance. We explore several issues of that type, e.g. the use of an initial heuristic or of a heuristic to solve subproblems arising at some step of an exact method, ordering of variables and branching rules as well as the trade-off between precision of bounds and size or difficulty of the programs used to compute them.