Mixed-Integer Programming Techniques for the Minimum Sum-of-Squares Clustering Problem

Martin Schmidt

@schmaidt

Séminaire du GERAD | February 24, 2022

The people who really did the work ...

Jan Pablo Burgard | Carina Moreira Costa



Christopher Hojny | Thomas Kleinert

Outlook

Motivation & Introduction

What is the MSSC problem?

Given

- n data points in $\mathbb{R}^d$: $p \in P \subset \mathbb{R}^d$, $|P| = n$
- $2 \leq k \leq n$ clusters

What is the MSSC problem?

Given

- n data points in $\mathbb{R}^d$: $p \in P \subset \mathbb{R}^d$, $|P| = n$
- $2 \leq k \leq n$ clusters

Task

Find ...

- an assignment for every data point to exactly one cluster
- and a representative for every cluster

What is the MSSC problem?

Given

- n data points in $\mathbb{R}^d$: $p \in P \subset \mathbb{R}^d$, $|P| = n$
- $2 \leq k \leq n$ clusters

Task

Find ...

- an assignment for every data point to exactly one cluster
- and a representative for every cluster

Key problem in unsupervised machine learning

A little bit more formal ...

Find k centroids c^j, $j = 1, \ldots, k$, that solve

$$\min_c \; h(P, c), \quad c = (c^j)_{j=1,\ldots,k},$$

A little bit more formal ...

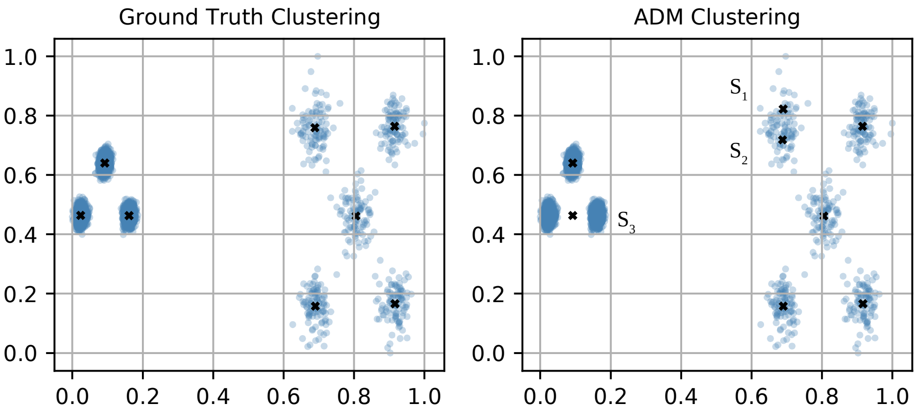
Find k centroids c^j, $j = 1, \ldots, k$, that solve

$$\min_c \; h(P, c), \quad c = (c^j)_{j=1,\ldots,k},$$

with

$$h(P, c) = \sum_{j=1}^{k} \sum_{p \in S_j} \|p - c^j\|_2^2$$

Test Instance Unbalance from Fränti and Sieranoja (2018)



Areas of Application

- The components of p are features of a measurement
  - gender
  - age
  - income
  - ...

Areas of Application

- The components of p are features of a measurement
  - gender
  - age
  - income
  - ...

Many Applications

- Data analysis: Cuesta-Albertos and Fraiman (2007), Sangalli et al. (2010)
- Market segmentation: Chen et al. (1998)
- Bio informatics: Datta and Datta (2003)
- Economics: He et al. (2007)
- Social science: Han (2022).

Hardness

The problem is NP-hard, even in two dimensions

- Aloise et al. (2009)
- Dasgupta (2007)
- Mahajan et al. (2012)

Hardness

The problem is NP-hard, even in two dimensions

- Aloise et al. (2009)
- Dasgupta (2007)
- Mahajan et al. (2012)

Many applications in practice only use heuristic approaches

- k-means clustering: MacQueen (1967), Lloyd (1982)

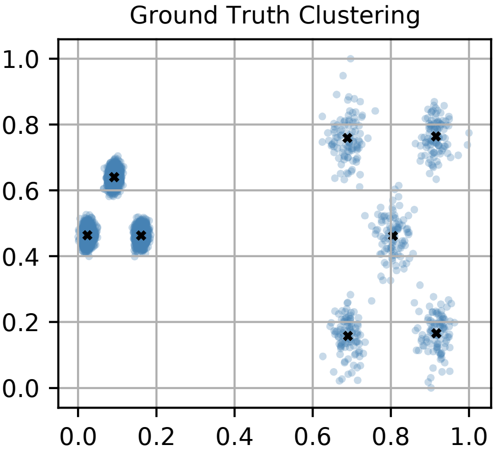
A Slide Stolen from Daniel Aloise

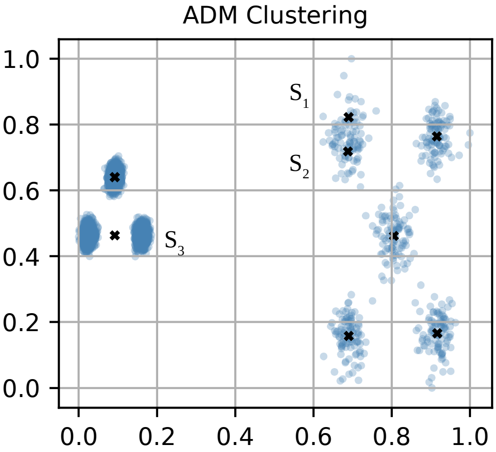
- *Convexity* is an important feature in optimization.
- Convexity means that global optimality can be obtained via local optimizers.
- Global optimality is particularly important for clustering.
  - As an unsupervised ML task, results usually require interpretation from domain experts.
  - Such interpretation can be completely erroneous in case the analyzed solution is far from the global optimum.
- Hence, a convex clustering model is a guarantee that the sole optimal solution is the best and the right one for posterior clustering analysis.

Global Optimality

- There are convex variants of the clustering problem
  - Convex fuzzy k-medoids clustering: Aloise et al. (2020)
- But: The MSSC problem is not convex!

- There are convex variants of the clustering problem
  - Convex fuzzy *k*-medoids clustering: Aloise et al. (2020)
- But: The MSSC problem is not convex!





A Short History of Solving the Problem to Global Optimality

- First branch-and-bound method: Fukunaga et al. (1975)
- Refinements: Diehr (1985) and Brusco (2006)
- Branch-and-bound with RLT: Sherali and Desai (2005) $\rightarrow$ Aloise and Hansen (2011)
- Symmetry breaking techniques: Plastria (2002), Sherali and Desai (2005)
- Column generation: Merle et al. (1999), Aloise et al. (2012)
- Benders: Floudas et al. (1989) and Tan et al. (2007)
- Conic reformulations: Prasad and Hanasusanto (2018)
- Voronoi diagrams: Tîrnăucă et al. (2018)
- SDP basics: Peng and Wei (2007) and Peng and Xia (2005b)
- SDP branch-and-cut: Aloise and Hansen (2009) $\rightarrow$ 202 data points
- SDP latest news: Piccialli et al. (2021) $\rightarrow$ 4000 data points
- Reduced-space techniques: Hua et al. (2021) and Liberti and Manca (2021)

Two MINLP Formulations

MINLP Modeling

- Introduce binary variables $x_{pj} \in \{0, 1\}$ for $p \in P$ and $j \in [k] := \{1, \dots, k\}$
- Reformulate the function h as

$$h(P, c, x) = \sum_{j \in [k]} \sum_{p \in P} x_{pj} \|p - c^j\|_2^2, \quad x = (x_{pj})_{p \in P}^{j \in [k]}$$

- Binary variables have the meaning

$$x_{pj} = \begin{cases} 1, & \text{if point } p \text{ is assigned to cluster } j \\ 0, & \text{otherwise} \end{cases}$$

- Every $p \in P$ should belong to exactly only one cluster

$$\sum_{j=1}^{k} x_{pj} = 1 \quad \text{for all} \quad p \in P$$

The Classic MINLP Model

$$
\begin{aligned}
\min_{x,c} \quad & \sum_{p \in P} \sum_{j \in [k]} x_{pj} \|p - c^j\|^2 \\
\text{s.t.} \quad & \sum_{j=1}^{k} x_{pj} = 1, \quad p \in P \\
& x_{pj} \in \{0, 1\}, \quad p \in P,\ j \in [k] \\
& c^j \in B, \quad j \in [k]
\end{aligned}
$$

- Mixed-integer linear constraints
- Nonlinear (cubic) and nonconvex objective function

$$
\begin{aligned}
\min_{x,c,\eta} \quad & \sum_{p \in P} \sum_{j \in [k]} \eta_{pj} \\
\text{s.t.} \quad & \eta_{pj} \geq \|p - c^j\|^2 - M_p(1 - x_{pj}), \quad p \in P,\ j \in [k] \\
& \sum_{j \in [k]} x_{pj} = 1, \quad p \in P \\
& x_{pj} \in \{0, 1\}, \quad p \in P,\ j \in [k] \\
& c^j \in B, \quad j \in [k] \\
& \eta_{pj} \geq 0, \quad p \in P,\ j \in [k]
\end{aligned}
$$

- M_p need to be chosen sufficiently large $\rightarrow$ doable
- Objective function is linear now
- New quadratic but convex (conic) constraints

Mixed-Integer Programming Techniques

Cardinality Cuts

- First used in Sherali and Desai (2005) as well as Aloise et al. (2012)
- Main insight: optimal solutions never have empty clusters
- Clusters can never contains more than $|P| - k + 1$ points

$$1 \leq \sum_{p \in P} x_{pj} \leq |P| - k + 1, \quad j \in [k]$$

Cardinality Cuts

- First used in Sherali and Desai (2005) as well as Aloise et al. (2012)
- Main insight: optimal solutions never have empty clusters
- Clusters can never contains more than $|P| - k + 1$ points

$$1 \leq \sum_{p \in P} x_{pj} \leq |P| - k + 1, \quad j \in [k]$$

These cuts can also be localized … but it didn't help much (more)

Outer Approximation Cuts

- Outer approximation is a powerful technique for solving convex MINLPs
  - Duran and Grossmann (1986), Fletcher and Leyffer (1994), Quesada and Grossmann (1992)
- The constraints

$$\eta_{pj} \geq \|p - c^j\|^2 - M_p(1 - x_{pj}), \quad p \in P,\ j \in [k],$$

  are nonlinear but convex
- 1st-order Taylor approximation at any point $(\bar{\eta}, \bar{c}, \bar{x})$ leads to global underestimators

$$\sum_{i=1}^{d} 2\bar{c}_i^j c_i^j - 2p_i c_i^j + (p_i)^2 - (\bar{c}_i^j)^2 - \eta_{pj} - M_p(1 - x_{pj}) \leq 0, \quad p \in P,\ j \in [k]$$

Propagation

- Suppose we are at a node of the branch-and-bound tree
- Branching decisions and further reductions
  - Some variables might have been fixed or their bounds have been tightened
- Aim of propagation
  - Find further variable fixings or bound tightenings that are valid at the current node

Propagation

- Suppose we are at a node of the branch-and-bound tree
- Branching decisions and further reductions
  - Some variables might have been fixed or their bounds have been tightened
- Aim of propagation
  - Find further variable fixings or bound tightenings that are valid at the current node

Challenge

Every assignment of x-variables that satisfies

$$\sum_{j \in [k]} x_{pj} = 1, \quad p \in P,$$

can be extended to a feasible point.

$\rightarrow$ Useful propagation mechanisms need to exclude assignments that cannot be optimal!

Barycenter Propagation

- $P_j \subseteq P$: set of all $p \in P$ such that x_{pj} is fixed to 1 at the current node
- $P'_j \subseteq P$: set of all p such that x_{pj} has not been fixed to 0 yet
- Note: $P_j \subseteq P'_j$
- $\underline{z}$ and $\bar{z}$: lower and upper bound on z at the current node
- For $\emptyset \neq Q \subseteq P$, the optimal choice for its centroid is the barycenter

$$\mathcal{C}(Q) := \frac{1}{|Q|} \sum_{p \in Q} p$$

- The corresponding loss function reads

$$\mathcal{D}(Q) = \sum_{p \in Q} \|p - \mathcal{C}(Q)\|^2$$

Barycenter Propagation for the Objective Function Values

- For sets $Q \subseteq Q' \subseteq P$, we have $\mathcal{D}(Q) \leq \mathcal{D}(Q')$
- Lower bound on the objective is given by $\sum_{j=1}^{k} \mathcal{D}(P_j)$

Barycenter Propagation for the Objective Function Values

- For sets $Q \subseteq Q' \subseteq P$, we have $\mathcal{D}(Q) \leq \mathcal{D}(Q')$
- Lower bound on the objective is given by $\sum_{j=1}^{k} \mathcal{D}(P_j)$

It's different for the epigraph reformulation

- Consider a pair $(p, j) \in P \times [k]$
- If $p \notin P'_j$ and $\underline{\eta}_{pj} > 0$ holds, the current node can be pruned
- If $p \notin P'_j$ and $\underline{\eta}_{pj} = 0$, we can fix η_{pj} to 0

Barycenter Propagation for the Centroids

- Consider $P'_j \setminus P_j = \{p^1, \dots, p^s\}$ such that $p_i^1 \leq p_i^2 \leq \cdots \leq p_i^s$
- For each $r \in [s]_0$ with $[s]_0 := [s] \cup \{0\}$, we compute $\gamma^{j,r} = \mathcal{C}(P_j \cup \{p^1, \dots, p^r\})$.

Barycenter Propagation for the Centroids

- Consider $P'_j \setminus P_j = \{p^1, \ldots, p^s\}$ such that $p_i^1 \leq p_i^2 \leq \cdots \leq p_i^s$
- For each $r \in [s]_0$ with $[s]_0 := [s] \cup \{0\}$, we compute $\gamma^{j,r} = \mathcal{C}(P_j \cup \{p^1, \ldots, p^r\})$.

Lemma

A valid lower bound on c_i^j is given by $\min_{r \in [s]_0} \gamma_i^{j,r}$.

- Consider $P_j' \setminus P_j = \{p^1, \ldots, p^s\}$ such that $p_i^1 \leq p_i^2 \leq \cdots \leq p_i^s$
- For each $r \in [s]_0$ with $[s]_0 := [s] \cup \{0\}$, we compute $\gamma^{j,r} = \mathcal{C}(P_j \cup \{p^1, \ldots, p^r\})$.

Lemma

A valid lower bound on c_i^j *is given by* $\min_{r \in [s]_0} \gamma_i^{j,r}$.

Lemma

A valid upper bound on c_i^j *is given by* $\max_{r \in [s]_0} \mathcal{C}(P_j \cup \{p^{s-r}, \ldots, p^s\})_i$.

Lemma

There exists an optimal solution of MSSC with clusters $P_1, \ldots, P_k$ such that for each $j \in [k]$, we have $\mathrm{conv}(P_j) \cap P = P_j$.

Lemma

There exists an optimal solution of MSSC with clusters $P_1, \ldots, P_k$ such that for each $j \in [k]$, we have $\mathbf{conv}(P_j) \cap P = P_j$.

The propagator

- Compute $\mathbf{conv}(P_j)$ for each $j \in [k]$
- Consider now a point $p \in P \cap \mathbf{conv}(P_j)$
- If $p \notin P_j'$ holds, we can prune the current node
- Otherwise, x_{pj} can be fixed to 1

Attention: $\mathbf{conv}(P_j)$ can have $\Omega(2^d)$ many facets!

Lemma

Let $P_1 \cup \cdots \cup P_k$ be a partition of the finite set $P \subseteq \mathbb{R}^d$. Suppose $\mathrm{conv}(P_j) \cap P = P_j$ holds for each $j \in [k]$. Then, for every $q \in P \setminus P_j$,

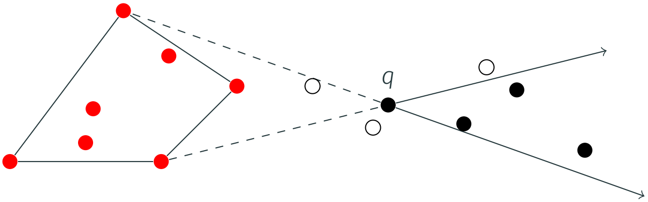
$$q + \mathrm{cone}\{-(p - q) \colon p \in P_j\} \subseteq P \setminus P_j.$$

Cone-Based Propagation

Lemma

Let $P_1 \cup \cdots \cup P_k$ be a partition of the finite set $P \subseteq \mathbb{R}^d$. Suppose $\mathbf{conv}(P_j) \cap P = P_j$ holds for each $j \in [k]$. Then, for every $q \in P \setminus P_j$,

$$q + \mathbf{cone}\{-(p - q) \colon p \in P_j\} \subseteq P \setminus P_j.$$



If q is not contained in the red cluster,
none of the black points can be contained in the red cluster.
Assigning the white points to the red cluster is still possible.

Distance Propagator

- Goal: Fix variables x_{pj}, $(p, j) \in P \times [k]$, to 0
- Centroid's bounding box $B_j = \{y \in \mathbb{R}^d : \underline{c}_i^j \leq y_i \leq \bar{c}_i^j,\ i \in [d]\}$
- Minimum/maximum distance of p to j

$$D_{j,p}^{\mathbf{min}} = \mathbf{min}\,\{\|p - x\| : x \in B_j\}\,, \quad D_{j,p}^{\mathbf{max}} = \mathbf{max}\,\{\|p - x\| : x \in B_j\}$$

- Easy to compute since B_j is a box
- Key insight: a point is assigned to a centroid of minimum distance in an optimal solution
- If there exists $j' \in [k]$ with $D_{j',p}^{\mathbf{max}} < D_{j,p}^{\mathbf{min}}$, x_{pj} can be set to 0

Branching Rules: General Ideas

- Suppose a node of the branch-and-bound tree has been solved
- Branching rules are used to split the problem so that two new subproblems are created
- Typically guided by the solution of the node's relaxation
- You want a primer on that topic? Achterberg et al. (2005)

Branching Rules: General Ideas

- Suppose a node of the branch-and-bound tree has been solved
- Branching rules are used to split the problem so that two new subproblems are created
- Typically guided by the solution of the node's relaxation
- You want a primer on that topic? Achterberg et al. (2005)

Intuition for what follows

- Interpret a fractional value x_{pj} as the probability that a point $p \in P$ is assigned to cluster $j \in [k]$
- Similar ideas but often (too) expensive: Gilpin and Sandholm (2011)

Entropy Branching

- Let the relaxed solution $\bar{x}$ be fractional
- Let $\bar{X}$ be the set of all branching candidates

$$\bar{X} := \{\bar{x}_{pj} : \bar{x}_{pj} \in (0,1),\ p \in P,\ j \in [k]\}$$

- Due to

$$\sum_{j=1}^{k} x_{pj} = 1, \quad p \in P,$$

  the value of $\bar{x}_{pj}$ can be seen as a probability of point p to belong to cluster j
- Idea: select the point p for which the probabilities for each cluster are almost the same
- Why? It reduces the “uncertainty” in the sub-tree

- Most unclear situation (homogeneous information setting)

$$\bar{x}_{pj} = \frac{1}{k} \quad \text{for all} \quad j \in [k]$$

- Let's use the Shannon entropy (Shannon, 1948)
- Consider the variables $\bar{x}_{pj} \in \bar{X}$
- Entropy of point p with probabilities $\bar{x}_{pj}$, $j \in [k]$, to be in cluster j:

$$H_p = -\sum_{j \in [k]} \bar{x}_{pj} \log_2(\bar{x}_{pj})$$

- Maximum entropy ($H_p = \log_2 k$) occurs in the case of homogeneous information
- Minimum entropy ($H_p = 0$) is obtained if there is a clear cluster assignment

- Find the point p^* corresponding to the fractional variable $\bar{x}_{p*j*} \in \bar{X}$ such that the entropy of point p^* is the maximal over all points with fractional variables
- In other words: we search for the most uncertain assignment
- Formally

$$p^* \in \underset{\{p \in P:\ \bar{x}_{pj} \in \bar{X},\, j \in [k]\}}{\arg\max} H_p$$

- For this point p^*, we branch on the first fractional variable $\bar{x}_{p*j}$

- Now a more geometric idea
- Rationale: clusters should be rather compact (opposed to being spread out)
- LP solution: centroid "suggestions" c^j, $j \in [k]$, and fractional variables $x_{pj} \in \bar{X}$
- Branching candidate: data point p and cluster j that are most apart from each other
- Formally:

$$(p^*, j^*) \in \underset{\{(p,j) \in P \times [k]:\ \bar{x}_{pj} \in \bar{X}\}}{\arg\max} \|p - c^j\|$$

Distance Branching

- Now a more geometric idea
- Rationale: clusters should be rather compact (opposed to being spread out)
- LP solution: centroid "suggestions" c^j, $j \in [k]$, and fractional variables $x_{pj} \in \bar{X}$
- Branching candidate: data point p and cluster j that are most apart from each other
- Formally:

$$(p^*, j^*) \in \underset{\{(p,j) \in P \times [k]:\ \bar{x}_{pj} \in \bar{X}\}}{\arg\max} \|p - c^j\|$$

- If an optimal cluster is indeed compact $\rightarrow$ the 0-subproblem contains an optimal solution
- Otherwise, the convexity propagator has the potential to fix additional variables that lie between p^* and the remaining points of cluster j^* in the 1-subproblem

Centrality Branching

- Idea of the distance branching rule
  - tailored towards the extremes of compact vs. far spread-out clusters
- Centrality branching rule
  - more balanced approach by selecting a point whose distance to a cluster is not too big
- Consider the LP solution with its suggestion for the centroids c^j, $j \in [k]$
- Idea: branch on the fractional variable x_{pj} corresponding to the data point p and cluster j that is lying in the center of the cloud of unassigned data points
- Cheap proxy: take the point p^* that is in the center of all centroids

$$p^* \in \underset{\{p \in P:\, \bar{x}_{pj} \in \bar{X},\, j \in [k]\}}{\arg\min} \sum_{j=1}^{k} \|p - c^j\|.$$

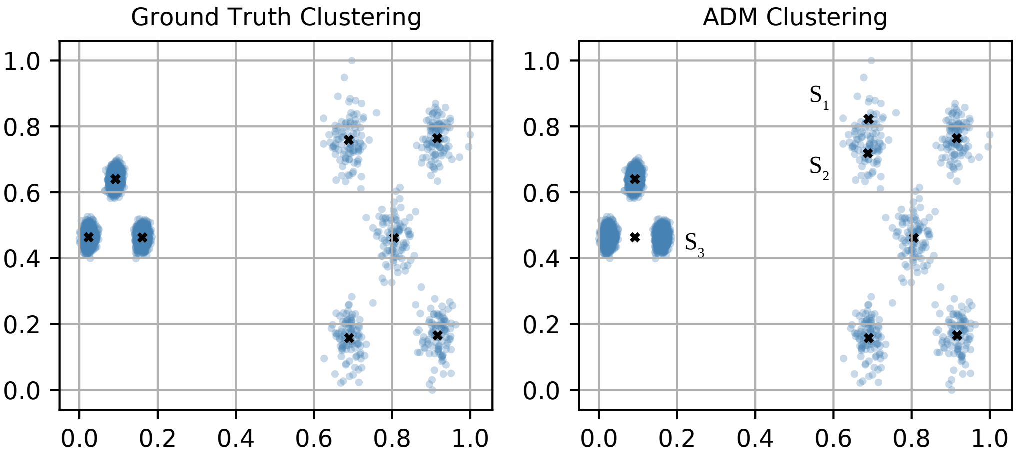
- For p^*, we branch on the first fractional variable $\bar{x}_{p^* j^*}$
- If the distance of p^* to cluster j^* is not too small, convexity propagator may fix further data points to be contained in cluster j^* in the 1-subproblem
- 0-subproblem: cone propagation might fix some variables to 0

A Root-Node Heuristic (that everybody knows)

- Well, just do what many practitioners do anyway: k-means clustering
- Lloyd (1982) and MacQueen (1967)
- Given: initial guess for the centroids
- Compute: assignment variables x
- Fix these assignments $\rightarrow$ compute new centroids
- … iterate …
- Initialization of first centroids with Maxmin heuristic
    - Gonzalez (1985) + Fränti and Sieranoja (2019)

A Rounding Heuristic

- Proposed first by Sherali and Desai (2005)
- Fractional LP solution $(\tilde{x}, \tilde{c})$
- Round the fractional $\tilde{x}$-solution to the closest feasible binary solution $\bar{x}$
- Respect the decisions that have already been made:
  if a data point is already assigned to a cluster, it will remain in that cluster
- Ensure no empty clusters: for each $j \in [k]$ with $P_j = \emptyset$, find the point $\bar{p} \in P \setminus \cup_{j=1}^{k} P_j$ such that $\bar{p} \in \operatorname{arg\,max} \left\{ \tilde{x}_{pj} : p \in P \setminus \cup_{j=1}^{k} P_j \right\}$
- Then fix $\bar{x}_{\bar{p}j} = 1$
- For each data point $p \in P$ with $\tilde{x}_{pj}$, $j \in [k]$, not yet rounded, find a cluster j^* such that $\tilde{x}_{pj*} \in \max \{\tilde{x}_{pj} : j \in [k]\}$
- Finally: compute centroids using the barycenter formula


Ground Truth Clustering
ADM Clustering
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
S1
S2
S3

An Improvement Heuristic

- Intra-Variance of cluster j as the weighted value of the loss function restricted to cluster C_j

$$F_j = \frac{1}{|C_j|} \sum_{p \in C_j} \|p - \bar{c}^j\|^2.$$

An Improvement Heuristic

- Intra-Variance of cluster j as the weighted value of the loss function restricted to cluster C_j

$$F_j = \frac{1}{|C_j|} \sum_{p \in C_j} \|p - \bar{c}^j\|^2.$$

- Joint centroid for a pair of clusters (C_{j_1}, C_{j_2})

$$c^{j_1 j_2} = \frac{1}{|C_{j_1}| + |C_{j_2}|} \sum_{p \in C_{j_1} \cup C_{j_2}} p$$

- Corresponding total loss

$$F_{j_1 j_2} = \frac{1}{|C_{j_1}| + |C_{j_2}|} \sum_{p \in C_{j_1} \cup C_{j_2}} \|p - c^{j_1 j_2}\|^2$$

An Improvement Heuristic

- Intra-Variance of cluster j as the weighted value of the loss function restricted to cluster C_j

$$F_j = \frac{1}{|C_j|} \sum_{p \in C_j} \|p - \bar{c}^j\|^2.$$

- Joint centroid for a pair of clusters (C_{j_1}, C_{j_2})

$$c^{j_1 j_2} = \frac{1}{|C_{j_1}| + |C_{j_2}|} \sum_{p \in C_{j_1} \cup C_{j_2}} p$$

- Corresponding total loss

$$F_{j_1 j_2} = \frac{1}{|C_{j_1}| + |C_{j_2}|} \sum_{p \in C_{j_1} \cup C_{j_2}} \|p - c^{j_1 j_2}\|^2$$

- Candidate triplets

$$\Psi := \{(C_{j_1}, C_{j_2}, C_{j_3}) : F_{j_1 j_2} < F_{j_3}\}$$

An Improvement Heuristic

- Candidate triplets

$$\Psi := \left\{ (C_{j_1}, C_{j_2}, C_{j_3}) \colon F_{j_1 j_2} < F_{j_3} \right\}$$

- $(C_{j_1}, C_{j_2}, C_{j_3}) \in \Psi$: total loss of joined clusters C_{j_1}, C_{j_2} is smaller than total loss within cluster C_{j_3}
- Idea: join clusters C_{j_1}, C_{j_2} and split C_{j_3}
- New centroids of splitted cluster

$$(\tilde{c}, \tilde{c}') \in \operatorname*{arg\,max}_{p, p' \in C_{j_3}} \left\{ \|p - p'\|^2 \right\}$$

Numerical Results

Software & Hardware Setup

- MINLP solver SCIP 7.0.3
- LP solver CPLEX 12.8
- All techniques are implemented as SCIP plugins in C/C++
- Intel Xeon CPU E5-2699 v4 at 2.20 GHz (2×44 threads) and 756 GB RAM
- Time limit: 1 h per instance
- Convex hull code: Qhull by Barber et al. (1996)

All codes and all data will be made available when the preprint is ready

Test Sets

Instance	Reference	n	d
Fisher150iris	Dua and Graff (2017) and Fisher (1936)	150	4
German22	Späth (1980)	22	2
German59	Späth (1980)	59	2
body-measurements	Heinz et al. (2003)	507	5
cities-coord-202	Grötschel (1991)	202	2
cities-coord-666	Grötschel (1991)	666	2
concrete-compressive	Dua and Graff (2017)	1030	8
glass-identification	Dua and Graff (2017)	214	9
image-segmentation	Dua and Graff (2017)	2310	19
padberg-rinaldi-hole-dri	Padberg and Rinaldi (1991)	2392	2
reinelt-hole-drilling	Reinelt (1991)	1060	2
ruspini	Ruspini (1970)	75	2
telugu-indian-vowel	Pal and Majumder (1977)	871	3
a1	Fränti and Sieranoja (2018)	3000	2
a2	Fränti and Sieranoja (2018)	5250	2
a3	Fränti and Sieranoja (2018)	7500	2
dim	Fränti and Sieranoja (2018)	1024	32
g2-2-30	Fränti and Sieranoja (2018)	2048	2
g2-2-50	Fränti and Sieranoja (2018)	2048	2
g2-2-70	Fränti and Sieranoja (2018)	2048	2
s1	Fränti and Sieranoja (2018)	5000	2
s2	Fränti and Sieranoja (2018)	5000	2
s3	Fränti and Sieranoja (2018)	5000	2
s4	Fränti and Sieranoja (2018)	5000	2
unbalance	Fränti and Sieranoja (2018)	6500	2

$k = 2$: Plain SCIP

Instance	n	d	Quad formulation			Epigraph formulation		
			nodes	time	gap	nodes	time	gap
Fisher150iris	150	4	909913	3600.01	∞	18956	3600.0	170.24
German22	22	2	207026	900.31	0.00	1343	17.25	0.00
German59	59	2	203438	3600.01	∞	112482	3600.0	32.98
body-measurements	507	5	38934	3600.0	∞	8862	3600.01	3338.16
cities-coord-202	202	2	84088	3600.0	∞	30276	3600.0	284.64
cities-coord-666	666	2	16459	3600.0	∞	19784	3600.01	2886.05
concrete-compressive	1030	8	12496	3600.05	∞	8505	3600.01	79149.66
glass-identification	214	9	240834	3600.0	∞	12176	3600.0	1920.57
image-segmentation	2310	19	3026	3600.0	∞	2870	3600.03	8405.89
padberg-rinaldi-hole-dri	2392	2	2654	3600.05	∞	114594	3600.02	∞
reinelt-hole-drilling	1060	2	2265	3600.01	∞	235206	3600.39	∞
ruspini	75	2	874673	3600.02	∞	137981	3600.01	132.28
telugu-indian-vowel	871	3	846	3600.2	∞	13681	3600.02	5082.69
a1	3000	2	3024	3600.0	∞	88111	3600.01	∞
a2	5250	2	5338	3600.01	∞	67159	3600.01	∞
a3	7500	2	6852	3600.01	∞	41941	3600.02	∞
dim	1024	32	3175	3600.01	∞	738	3600.28	∞
g2-2-30	2048	2	13927	3600.27	∞	10273	3600.02	22444.38
g2-2-50	2048	2	11804	3600.02	∞	10837	3600.01	13433.08
g2-2-70	2048	2	2376	3600.01	∞	8235	3600.01	19192.83
s1	5000	2	5032	3600.02	∞	58633	3600.0	∞
s2	5000	2	5050	3600.04	∞	47121	3600.0	∞
s3	5000	2	5035	3600.03	∞	55728	3600.03	∞
s4	5000	2	5029	3600.01	∞	53218	3600.01	∞
unbalance	6500	2	6544	3600.0	∞	37468	3600.86	∞

$k = 2$: SCIP + Heuristics + Convexity & Cone Propagator

Instance	n	d	Quad formulation			Epigraph formulation		
			nodes	time	gap	nodes	time	gap
Fisher150iris	150	4	1254550	3600.01	∞	38867	3600.0	114.71
German22	22	2	539	1.4	0.00	269	1.28	0.00
German59	59	2	2139	7.71	0.00	627	7.45	0.00
body-measurements	507	5	247712	3600.0	∞	10235	3600.0	1460.22
cities-coord-202	202	2	49373	142.42	0.00	4759	155.58	0.00
cities-coord-666	666	2	63323	773.03	0.00	3644	525.23	0.00
concrete-compressive	1030	8	962	3609.45	∞	6464	3600.0	31862.98
glass-identification	214	9	216426	3600.0	∞	12824	3600.0	877.58
image-segmentation	2310	19	2175	3600.0	∞	2193	3600.0	28016.73
padberg-rinaldi-hole-dri	2392	2	53	3600.03	∞	44395	3600.01	∞
reinelt-hole-drilling	1060	2	26	3600.06	∞	161496	3600.37	∞
ruspini	75	2	6875	15.21	0.00	549	10.29	0.00
telugu-indian-vowel	871	3	361	3600.18	∞	8627	3600.01	171.65
a1	3000	2	340	3600.07	∞	40987	3600.0	∞
a2	5250	2	22	3600.02	∞	10912	3600.0	∞
a3	7500	2	219	3600.07	∞	5419	3600.0	∞
dim	1024	32	3318	3600.03	∞	1655	3600.01	∞
g2-2-30	2048	2	172536	3600.0	∞	6249	3460.79	0.00
g2-2-50	2048	2	205102	3600.0	∞	4509	3600.03	11.25
g2-2-70	2048	2	190472	3600.0	∞	3714	3600.0	44.85
s1	5000	2	206	3600.03	∞	15937	3600.07	∞
s2	5000	2	58	3600.1	∞	14702	3600.0	∞
s3	5000	2	70	3600.01	∞	14203	3600.05	∞
s4	5000	2	34	3600.16	∞	16303	3600.0	∞
unbalance	6500	2	24	3600.01	∞	20735	3600.8	∞

$k = 2$: SCIP + Heuristics + Barycenter Propagator

Instance	n	d	Quad formulation			Epigraph formulation		
			nodes	time	gap	nodes	time	gap
Fisher150iris	150	4	1291451	3600.06	390.75	36671	3600.0	133.79
German22	22	2	735	1.0	0.00	603	3.14	0.00
German59	59	2	298474	3600.0	60.67	181024	3302.02	0.00
body-measurements	507	5	175149	3600.0	4020.08	10707	3600.0	2023.87
cities-coord-202	202	2	209526	3600.0	182.11	53429	3600.0	135.06
cities-coord-666	666	2	6326	3600.0	2030.10	29584	3600.0	1624.17
concrete-compressive	1030	8	78715	3600.03	5124.69	5448	3600.01	18755.18
glass-identification	214	9	378677	3600.01	327.75	13634	3600.0	449.85
image-segmentation	2310	19	13938	3600.03	41771.35	251	3600.0	∞
padberg-rinaldi-hole-dri	2392	2	215038	3600.01	5111.74	115222	3600.01	∞
reinelt-hole-drilling	1060	2	548760	3600.01	1821.46	243529	3600.03	∞
ruspini	75	2	371336	3600.01	172.24	202409	3600.0	62.62
telugu-indian-vowel	871	3	482986	3600.01	1188.41	11452	3600.0	2802.01
a1	3000	2	154455	3600.01	9892.16	76186	3600.02	∞
a2	5250	2	79802	3600.01	14609.94	28009	3600.01	∞
a3	7500	2	50747	3600.02	28536.14	8957	3600.02	∞
dim	1024	32	20279	3600.0	12806.54	938	3600.01	∞
g2-2-30	2048	2	68620	3600.0	10686.30	9677	3600.01	8018.12
g2-2-50	2048	2	69035	3600.01	8460.72	11622	3600.01	6694.39
g2-2-70	2048	2	75071	3600.0	6784.35	10224	3600.0	6502.58
s1	5000	2	78784	3600.02	29879.42	30042	3600.13	∞
s2	5000	2	82041	3600.01	18242.68	34579	3600.0	∞
s3	5000	2	87351	3600.0	17978.65	32344	3600.01	∞
s4	5000	2	87945	3600.0	17583.74	35208	3600.36	∞
unbalance	6500	2	70452	3600.01	5680.62	22020	3600.62	∞

$k = 2$: SCIP + Heuristics + Convexity & Cone & Barycenter Propagator

Instance	n	d	Quad formulation			Epigraph formulation		
			nodes	time	gap	nodes	time	gap
Fisher150iris	150	4	1333676	3600.01	180.59	47822	3600.0	84.02
German22	22	2	155	0.21	0.00	179	0.96	0.00
German59	59	2	509	0.98	0.00	631	7.57	0.00
body-measurements	507	5	316874	3600.0	3781.81	12521	3600.0	1281.91
cities-coord-202	202	2	3385	5.74	0.00	3954	139.4	0.00
cities-coord-666	666	2	2800	14.04	0.00	4116	562.66	0.00
concrete-compressive	1030	8	91207	3600.01	5087.22	5158	3600.01	18171.42
glass-identification	214	9	542435	3600.01	326.10	13759	3600.0	442.45
image-segmentation	2310	19	17008	3600.01	39499.81	257	3600.03	∞
padberg-rinaldi-hole-dri	2392	2	1469	3600.0	23.52	80303	3600.17	∞
reinelt-hole-drilling	1060	2	1788	3600.1	19.02	155237	3600.04	∞
ruspini	75	2	487	0.95	0.00	505	9.96	0.00
telugu-indian-vowel	871	3	28388	3600.15	60.12	7274	3600.01	143.63
a1	3000	2	1718	3600.05	25.16	35274	3600.0	∞
a2	5250	2	936	3600.06	30.58	13296	3600.04	∞
a3	7500	2	886	3600.0	84.06	6954	3600.01	∞
dim	1024	32	21686	3600.05	12465.58	884	3600.01	∞
g2-2-30	2048	2	6383	114.73	0.00	5959	3600.02	7.10
g2-2-50	2048	2	14995	217.03	0.00	5967	3600.01	13.32
g2-2-70	2048	2	36329	370.0	0.00	2531	3600.01	19.68
s1	5000	2	2367	3600.07	244.12	13093	3600.01	∞
s2	5000	2	676	3600.01	173.69	15993	3600.0	∞
s3	5000	2	3971	3600.01	133.13	11323	3600.22	∞
s4	5000	2	4708	3600.02	102.62	23422	3600.01	∞
unbalance	6500	2	181	3600.02	∞	15874	3600.0	∞

$k = 2$: SCIP + Heuristics + Convexity & Cone & Barycenter Propagator + OA

Instance	n	d	Quad formulation			Epigraph formulation		
			nodes	time	gap	nodes	time	gap
Fisher150iris	150	4	1333676	3600.01	180.59	62669	3600.0	63.93
German22	22	2	155	0.21	0.00	247	1.29	0.00
German59	59	2	509	0.98	0.00	571	7.03	0.00
body-measurements	507	5	316874	3600.0	3781.81	11111	3600.03	1751.01
cities-coord-202	202	2	3385	5.74	0.00	4146	158.78	0.00
cities-coord-666	666	2	2800	14.04	0.00	5023	780.78	0.00
concrete-compressive	1030	8	91207	3600.01	5087.22	4321	3600.01	14466.75
glass-identification	214	9	542435	3600.01	326.10	14631	3600.0	305.74
image-segmentation	2310	19	17008	3600.01	39499.81	904	3600.03	10007.79
padberg-rinaldi-hole-dri	2392	2	1469	3600.0	23.52	481	3600.01	2079.74
reinelt-hole-drilling	1060	2	1788	3600.1	19.02	606	3600.03	250.56
ruspini	75	2	487	0.95	0.00	573	9.47	0.00
telugu-indian-vowel	871	3	28388	3600.15	60.12	4932	3600.0	147.79
a1	3000	2	1718	3600.05	25.16	3645	3600.0	754130.50
a2	5250	2	936	3600.06	30.58	691	3600.02	16336991.23
a3	7500	2	886	3600.0	84.06	48	3600.07	26935734.14
dim	1024	32	21686	3600.05	12465.58	452	3600.01	∞
g2-2-30	2048	2	6383	114.73	0.00	6825	3600.02	3.23
g2-2-50	2048	2	14995	217.03	0.00	5724	3600.01	16.97
g2-2-70	2048	2	36329	370.0	0.00	4131	3600.01	29.80
s1	5000	2	2367	3600.07	244.12	176	3600.07	1227922.26
s2	5000	2	676	3600.01	173.69	180	3600.04	2062192.77
s3	5000	2	3971	3600.01	133.13	159	3600.09	59440.31
s4	5000	2	4708	3600.02	102.62	1352	3600.06	12472.41
unbalance	6500	2	181	3600.02	∞	394	3600.11	336156.23

$k = 2$: SCIP + Heuristics + All Propagators + OA

Instance	n	d	Quad formulation			Epigraph formulation		
			nodes	time	gap	nodes	time	gap
Fisher150iris	150	4	1594034	3600.01	163.35	89287	3600.0	53.12
German22	22	2	155	0.18	0.00	247	1.12	0.00
German59	59	2	509	0.83	0.00	571	5.63	0.00
body-measurements	507	5	374487	3600.01	3670.49	14984	3600.0	1609.83
cities-coord-202	202	2	3383	5.4	0.00	4146	111.59	0.00
cities-coord-666	666	2	2800	13.41	0.00	5023	648.54	0.00
concrete-compressive	1030	8	99409	3600.01	5038.68	5290	3600.01	14466.75
glass-identification	214	9	653118	3600.01	326.46	20036	3600.0	286.45
image-segmentation	2310	19	18400	3600.0	38483.96	1200	3600.02	10007.79
padberg-rinaldi-hole-dri	2392	2	1469	3600.02	23.52	598	3600.01	882.60
reinelt-hole-drilling	1060	2	1788	3600.03	19.02	625	3600.02	250.56
ruspini	75	2	487	0.92	0.00	573	7.89	0.00
telugu-indian-vowel	871	3	28388	3600.15	60.12	6074	3600.0	132.42
a1	3000	2	1718	3600.04	25.16	7080	3600.0	754130.50
a2	5250	2	936	3600.01	30.58	828	3600.0	3488806.65
a3	7500	2	886	3600.05	84.06	49	3600.04	26935734.14
dim	1024	32	24323	3600.03	12218.34	573	3600.0	∞
g2-2-30	2048	2	6383	87.79	0.00	6989	2924.55	0.00
g2-2-50	2048	2	14995	172.4	0.00	7800	3600.01	11.49
g2-2-70	2048	2	36329	320.03	0.00	5514	3600.02	19.66
s1	5000	2	2367	3600.0	244.12	183	3600.08	1124074.89
s2	5000	2	676	3600.02	173.69	211	3600.04	2062192.77
s3	5000	2	2815	3600.01	220.62	213	3600.13	12046.53
s4	5000	2	2657	3600.01	182.74	1512	3600.02	11820.60
unbalance	6500	2	181	3600.01	∞	461	3600.06	269814.70

$k = 2$: SCIP + Heuristics + All Propagators + OA + Entropy Branching

Instance	n	d	Quad formulation			Epigraph formulation		
			nodes	time	gap	nodes	time	gap
Fisher150iris	150	4	1462878	3600.01	226.87	68847	3600.0	169.43
German22	22	2	357	0.43	0.00	155	0.69	0.00
German59	59	2	765	1.32	0.00	649	4.26	0.00
body-measurements	507	5	410938	3600.0	3632.71	18072	3600.0	1396.68
cities-coord-202	202	2	3541	5.47	0.00	3249	68.49	0.00
cities-coord-666	666	2	2604	11.45	0.00	5308	690.09	0.00
concrete-compressive	1030	8	104887	3600.0	5502.52	10227	3600.0	10001.54
glass-identification	214	9	650169	3600.07	567.57	26316	3600.0	304.25
image-segmentation	2310	19	22334	3600.02	51123.06	885	3600.03	9677.16
padberg-rinaldi-hole-dri	2392	2	1080	3600.0	32.60	748	3600.0	56479.73
reinelt-hole-drilling	1060	2	1857	3600.08	16.07	1243	3600.01	686.79
ruspini	75	2	1039	1.81	0.00	555	5.95	0.00
telugu-indian-vowel	871	3	160309	580.64	0.00	13930	3600.0	695.29
a1	3000	2	2863	3600.03	18.17	3496	3600.01	177852.67
a2	5250	2	936	3600.03	30.58	1677	3600.0	26840744.46
a3	7500	2	886	3600.03	84.06	1604	3600.0	1784421.45
dim	1024	32	11684	3600.0	12621.01	1488	3600.02	∞
g2-2-30	2048	2	6123	79.84	0.00	6399	2314.59	0.00
g2-2-50	2048	2	8169	3600.04	6.84	9044	3600.04	15.52
g2-2-70	2048	2	39249	455.03	0.00	6310	3600.03	32.64
s1	5000	2	2367	3600.0	244.12	1690	3600.02	91869.90
s2	5000	2	676	3600.02	173.69	433	3600.0	163413.35
s3	5000	2	2815	3600.01	220.62	173	3600.07	71533.78
s4	5000	2	2657	3600.01	182.74	2898	3600.0	600335.95
unbalance	6500	2	181	3600.02	∞	459	3600.05	179555.61

That's (Almost) It

- We get from nothing to
  - 8/25 solved instances
  - All other instances have a finite gap
- Solving the problem to global optimality still stays an extremely challenging task
- We can solve "moderate-sized" instances
  - Idea: Combine with reduction-techniques
  - Reduced-space branch-and-bound method by Hua et al. (2021)
  - Random projections by Liberti and Manca (2021)

That's (Almost) It

- We get from nothing to
  - 8/25 solved instances
  - All other instances have a finite gap
- Solving the problem to global optimality still stays an extremely challenging task
- We can solve "moderate-sized" instances
  - Idea: Combine with reduction-techniques
  - Reduced-space branch-and-bound method by Hua et al. (2021)
  - Random projections by Liberti and Manca (2021)

Preprint can be found soon at ...

SCAN ME