# $0-1$ SDP Approach to the Minimum Sum-of-Squares Clustering

Jiming Peng, Franz Rendl, and Yu Xia

Large Scale Nonlinear and Semidefinite Programming
University of Waterloo
May 2004

# Minimum Sum of Squares Clustering

- **Problem Definition**
  - Given a set of entities, partition them into clusters, such that the sum of squared distances from each entity to the centroid is minimum.
- **Applications**
  - Facilities location
  - Bioinformatics
  - Pattern recognition
  - Text retrieval, etc.

# Mathematical Model

Given $n$ points $\mathbf{a}_i \in \mathbb{R}^d$, find $k$ centroids $\mathbf{c}_i \in \mathbb{R}^d$:

$$\min_{\mathbf{c}_1,\ldots,\mathbf{c}_k} \sum_{i=1}^{n} \min \left\{ \|\mathbf{a}_i - \mathbf{c}_1\|_2^2, \ldots, \|\mathbf{a}_i - \mathbf{c}_k\|_2^2 \right\}.$$

**K-Means Algorithm**

Reassign one entity at a time until stability is reached.

1. Place k entities as initial group centroids.
2. Assign each object to the group that has the closest centroid.
3. When all objects have been assigned, recalculate the centroids.
4. Repeat Steps 2 and 3 until the centroids no longer move.

## The $0-1$ Nonlinear Programming Formulation

$$x_{ij} = \begin{cases} 1 & \mathbf{a}_i \text{ assigned to } \mathbf{c}_j; \\ 0 & \text{otherwise.} \end{cases}$$

$$\begin{aligned} \min_{x_{ij}} & \sum_{j=1}^{k}\sum_{i=1}^{n} x_{ij} \left\| \mathbf{a}_i - \frac{\sum_{i=1}^{n} x_{ij}\mathbf{a}_i}{\sum_{i=1}^{n} x_{ij}} \right\|_2^2 \\ \text{s.t.} & \sum_{j=1}^{k} x_{ij} = 1 \ (i = 1, \ldots, n) \\ & x_{ij} \in \{0, 1\} \ (i = 1, \ldots, n;\ j = 1, \ldots, k) \end{aligned} \quad (1)$$

## Difficulties:

- Nonlinear and nonconvex object.
- Discrete variables.
- Hard to design approximation method.

## New Formulation Based on SDP

- Let $X = [x_{ij}] \in \mathbb{R}^{n\times k}$ be the assignment matrix. Let $Z = X(X^TX)^{-1}X^T$
- The objective function is

$$\operatorname{tr} A^TA - \operatorname{tr}(A^TAZ).$$

- $Z$ is a projection matrix, $Z^2 = Z$, $z_{ij} \geq 0$.
- 

$$X\mathbf{e} = \mathbf{e} \Longrightarrow Z\mathbf{e} = \mathbf{e}.$$

- 

$$\operatorname{tr} Z = k.$$

# The $0-1$ SDP Model

(0-1 SDP)

$$
\begin{aligned}
\min \quad & \operatorname{tr} AA^T(I-Z) \\
\text{s.t.} \quad & Z\mathbf{e} = \mathbf{e} \\
& \operatorname{tr} Z = k \\
& Z \geq 0 \\
& Z^2 = Z
\end{aligned}
$$

**Theorem** The $0-1$ SDP model gives the exact solution.

## Difficulties for the SDP Model

Nonlinear SDP constraint:

$$Z^2 = Z.$$

## Solution

- K-Means-type algorithm can be viewed as a special heuristic method for it.
- Relax the SDP constraint.
- Approximate the feasible region by linear constraints.

## The LP Approximation

Define

$$\text{MET} = \{z_{ij} \le z_{ii},\ z_{ij} \le z_{jj},\ z_{ij} + z_{il} \le z_{ii} + z_{jl}\}.$$

(LP-MET)

$$\begin{aligned} \min \quad & \operatorname{tr} AA^T(I - Z) \\ \text{s.t.} \quad & Z\mathbf{e} = \mathbf{e} \\ & \operatorname{tr} Z = k \\ & Z \ge 0 \\ & Z \in \text{MET} \end{aligned}$$

# Numerical Examples based on the LP Model

This is the first time global optimum obtained without enumerating all the vertices.

**The Algorithm**

1. Solve the LP-MET model.
2. Find the pattern of $Z$ as $P^T DP : P$ permutation matrix, $D$ block diagonal matrix with each block has equal entries: $\frac{1}{n_i}$, ($n_i$: the size of the block).
3. Use K-Means to reassign some fuzzy points.

**Datasets**

- The Soybean data from UCI. $d = 35$, $n = 47$.
- The Ruspini data. $d = 2$, $n = 75$.
- The Späth's postal zones data. $d = 3$, $n = 89$.

# Results

The Ruspini's data.

| k | Objective | CPU time(s) |
|---|---|---|
| 2 | 893380 | 27.81 |
| 3 | 510630 | 66.58 |
| 4 | 12881 | 7.22 |
| 5 | 10127 | 9.47 |

The Soybean data.

| k | Objective | CPU time(s) |
|---|---|---|
| 2 | 404.4593 | 4.26 |
| 3 | 215.2593 | 1.51 |
| 4 | 205.9637 | 1.68 |

The Spath's Postal Zone data.

| k | Objective | CPU time(s) |
|---|---|---|
| 2 | $6.0255 * 10^{11}$ | 283.26 |
| 3 | $2.9451 * 10^{11}$ | 418.07 |
| 4 | $1.0447 * 10^{11}$ | 99.54 |
| 5 | $5.9761 * 10^{11}$ | 60.67 |
| 6 | $3.5908 * 10^{11}$ | 52.55 |
| 7 | $2.1983 * 10^{11}$ | 61.78 |
| 8 | $1.3385 * 10^{11}$ | 26.91 |
| 9 | $7.8044 * 10^{11}$ | 18.04 |

# Challenges and Future Work

- **Challenges**
  The LP-MET model can only solve small size problems.
- **Future Work**
  Use Lagrangian Dual and Bundle method for the LP apporximation model.