Partially Optimal Cubic Subspace Clustering

Research Project Machine Learning

Volodymyr Drobitko

Technische Universität Dresden

21.07.2025

Problem Statement (1)

Finite sample set S, cost function $\mathrm{c}\colon \binom{S}{3} \to \mathbb{R}$.
Instance of the **Cubic Clique Partition Problem**:

$$\max_{\mathrm{y}\colon \binom{S}{2} \to \{0,1\}} \sum_{\{a,b,c\} \in \binom{S}{3}} \mathrm{c}_{\{a,b,c\}}\, \mathrm{y}_{\{a,b\}}\, \mathrm{y}_{\{b,c\}}\, \mathrm{y}_{\{a,c\}}$$

subject to $\mathrm{y}_{\{a,b\}} + \mathrm{y}_{\{b,c\}} - 1 \leq \mathrm{y}_{\{a,c\}}$ for all distinct $a, b, c \in S$.

Problem Statement (1)

Finite sample set S, cost function $\mathrm{c}\colon \binom{S}{3} \to \mathbb{R}$.
Instance of the **Cubic Clique Partition Problem**:

$$\max_{\mathrm{y}\colon \binom{S}{2} \to \{0,1\}} \sum_{\{a,b,c\} \in \binom{S}{3}} \mathrm{c}_{\{a,b,c\}}\, \mathrm{y}_{\{a,b\}}\, \mathrm{y}_{\{b,c\}}\, \mathrm{y}_{\{a,c\}}$$

subject to $\mathrm{y}_{\{a,b\}} + \mathrm{y}_{\{b,c\}} - 1 \leq \mathrm{y}_{\{a,c\}}$ for all distinct $a, b, c \in S$.

Find a **partially optimal solution**, i.e. fix some labels $\mathrm{y}_{\{a,b\}}$ for distinct $a, b \in S$

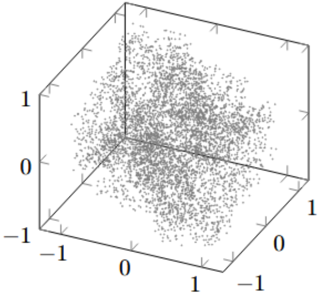
$$\begin{cases} \mathrm{y}_{\{a,b\}} = 1 & \text{join } a, b \\ \mathrm{y}_{\{a,b\}} = 0 & \text{cut } a, b \\ \mathrm{y}_{\{a,b\}} = ? & \text{unknown} \end{cases}$$

in such way that there still exists an optimal solution.

Problem Statement (2)

Subspace Instances of the Cubic Clique Partition Problem

Samples S: points $S \subset \mathbb{R}^3$

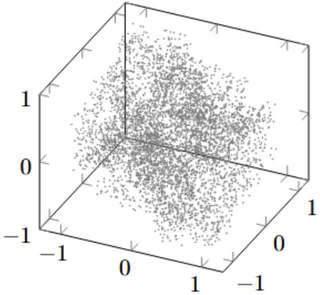


(a) Samples S

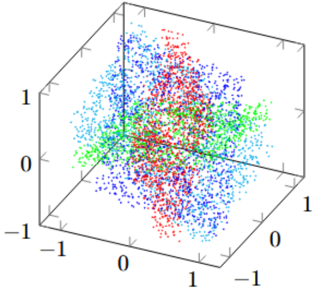
Problem Statement (2)

Subspace Instances of the Cubic Clique Partition Problem

Samples S: points $S \subset \mathbb{R}^3$
Point generation: 3 distinct planes containing the origin, noise σ



(a) Samples S

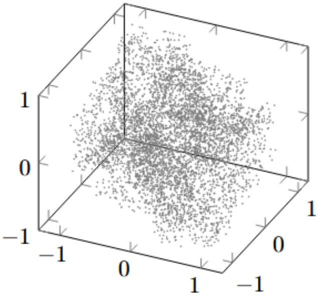


(b) Optimal clustering y^*

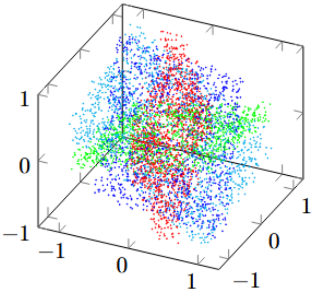
Problem Statement (2)

Subspace Instances of the Cubic Clique Partition Problem

Samples S: points $S \subset \mathbb{R}^3$
Point generation: 3 distinct planes containing the origin, noise σ
Optimal clustering y*: original planes



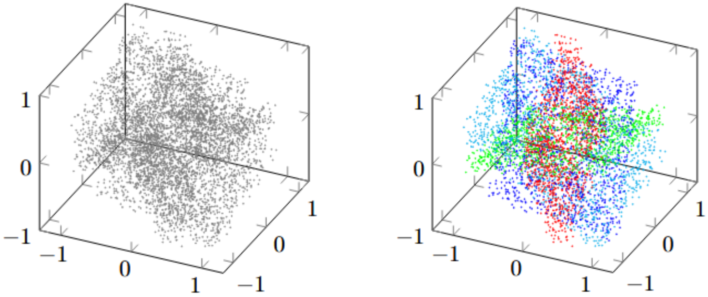
(a) Samples S



(b) Optimal clustering y*

Subspace Instances of the Cubic Clique Partition Problem

Samples S: points $S \subset \mathbb{R}^3$
Point generation: 3 distinct planes containing the origin, noise σ
Optimal clustering y^*: original planes
Cost function c? (no concrete plane info given)



(a) Samples S

(b) Optimal clustering y^*

Research Goals and Contributions

Task	Solution
read 'Partial Optimality in Cubic Correlation Clustering' [?]; implement the algorithm for establishing partial optimality to the cubic clique partition problem;	

Research Goals and Contributions

Task	**Solution**
read 'Partial Optimality in Cubic Correlation Clustering' [**?**]; implement the algorithm for establishing partial optimality to the cubic clique partition problem;	my own implementation in C++ of the suggested algorithm with some adjustments

Research Goals and Contributions

Task	**Solution**
read 'Partial Optimality in Cubic Correlation Clustering' [**?**]; implement the algorithm for establishing partial optimality to the cubic clique partition problem;	my own implementation in C++ of the suggested algorithm with some adjustments
construct subspace instances of increasing difficulty by generating the sample points and by defining a **suitable cost function** for the point triples	

Research Goals and Contributions

Task	Solution
read 'Partial Optimality in Cubic Correlation Clustering' [**?**]; implement the algorithm for establishing partial optimality to the cubic clique partition problem;	my own implementation in C++ of the suggested algorithm with some adjustments
construct subspace instances of increasing difficulty by generating the sample points and by defining a **suitable cost function** for the point triples	point generation using the linear algebra methods; experimentally determined geometric cost function with a significant noise tolerance;

Research Goals and Contributions

Task	**Solution**
read 'Partial Optimality in Cubic Correlation Clustering' [**?**]; implement the algorithm for establishing partial optimality to the cubic clique partition problem;	my own implementation in C++ of the suggested algorithm with some adjustments
construct subspace instances of increasing difficulty by generating the sample points and by defining a **suitable cost function** for the point triples	point generation using the linear algebra methods; experimentally determined geometric cost function with a significant noise tolerance;
apply the algorithm to the constructed subspace instances, assess partial optimality, accuracy with respect to truth, and computation time	

Research Goals and Contributions

Task	Solution
read 'Partial Optimality in Cubic Correlation Clustering' [**?**]; implement the algorithm for establishing partial optimality to the cubic clique partition problem;	my own implementation in C++ of the suggested algorithm with some adjustments
construct subspace instances of increasing difficulty by generating the sample points and by defining a **suitable cost function** for the point triples	point generation using the linear algebra methods; experimentally determined geometric cost function with a significant noise tolerance;
apply the algorithm to the constructed subspace instances, assess partial optimality, accuracy with respect to truth, and computation time	systematic empirical assessment of the algorithm application the the problem subspace instances with my cost function proving its quality