Partially Optimal Cubic Subspace Clustering

Research Project Machine Learning

Volodymyr Drobitko

Technische Universität Dresden

21.07.2025

Introduction

2 Partial Optimality for Cubic Clique Partition Problem

3 Cubic Subspace Instance Construction

Experiments and Evaluation

Conclusion

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

$$\min_{\mathbf{y}: \; \binom{S}{2} \to \{0,1\}} \sum_{abc \in \binom{S}{3}} c_{abc} \, \mathbf{y}_{ab} \, \mathbf{y}_{bc} \, \mathbf{y}_{ac}$$

subject to $\mathbf{y}_{ab} + \mathbf{y}_{bc} - 1 \leq \mathbf{y}_{ac}$ for all distinct $a,b,c \in \mathcal{S}$.

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

$$\min_{\mathbf{y}: \, \binom{S}{2} \rightarrow \{0,1\}} \sum_{abc \in \binom{S}{3}} \mathbf{c}_{abc} \, \mathbf{y}_{ab} \, \mathbf{y}_{bc} \, \mathbf{y}_{ac}$$

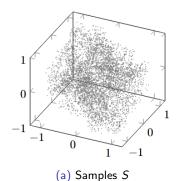
subject to $y_{ab} + y_{bc} - 1 \le y_{ac}$ for all distinct $a, b, c \in S$.

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

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

in such way that there still exists an optimal solution.

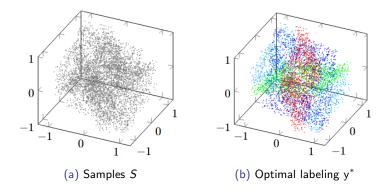
Subspace Instances of the Cubic Clique Partition Problem Samples S: points $S \subset \mathbb{R}^3$



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

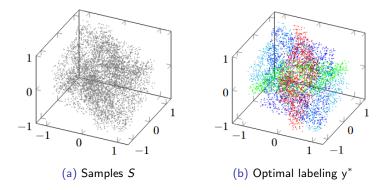


Subspace Instances of the Cubic Clique Partition Problem

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

Point generation: 3 distinct planes containing the origin, noise σ

Optimal labeling y*: original planes



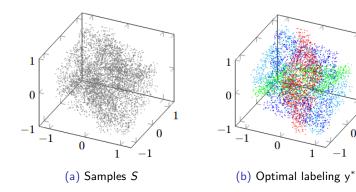
Subspace Instances of the Cubic Clique Partition Problem

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

Point generation: 3 distinct planes containing the origin, noise σ

Optimal labeling y*: original planes

Cost function c? (no concrete plane information given)



Related Work:

TODO: 2 + 1

Related Work:

TODO: 2 + 1

Tasks and Solutions:

■ Read TODO, implement the partial optimality algorithm

 → implementation in C++ (with some adjustments)

Related Work:

TODO: 2 + 1

Tasks and Solutions:

- Read TODO, implement the partial optimality algorithm \rightarrow implementation in C++ (with some adjustments)
- ② Construct subspace instances of increasing difficulty
 → point generation, appropriate cost function c
 (significant noise tolerance)

Related Work:

TODO: 2 + 1

Tasks and Solutions:

- Read TODO, implement the partial optimality algorithm \rightarrow implementation in C++ (with some adjustments)
- ② Construct subspace instances of increasing difficulty
 → point generation, appropriate cost function c (significant noise tolerance)
- Apply algorithm to the subspace instances, assess partial optimality, accuracy and computation time
 - \rightarrow experiments and evaluation (prove the quality of c)

Introduction

Partial Optimality for Cubic Clique Partition Problem

3 Cubic Subspace Instance Construction

4 Experiments and Evaluation

Conclusion

Extended cost function c: $\binom{S}{3} \cup \binom{S}{2} \cup \emptyset \rightarrow \mathbb{R}$

Extended cost function c: $\binom{S}{3} \cup \binom{S}{2} \cup \emptyset \rightarrow \mathbb{R}$

Instance of the extended cubic clique partition problem:

$$\min_{\mathtt{y}: \, \binom{S}{2} \to \{0,1\}} \sum_{abc \in \binom{S}{3}} \mathsf{c}_{abc} \, \mathtt{y}_{ab} \, \mathtt{y}_{bc} \, \mathtt{y}_{ac} + \sum_{ab \in \binom{S}{2}} \mathsf{c}_{ab} \, \mathtt{y}_{ab} + \mathsf{c}_{\emptyset}$$

subject to $\mathbf{y}_{ab} + \mathbf{y}_{bc} - 1 \leq \mathbf{y}_{ac}$ for all distinct $a,b,c \in \mathcal{S}$.

Extended cost function c: $\binom{S}{3} \cup \binom{S}{2} \cup \emptyset \rightarrow \mathbb{R}$

Instance of the extended cubic clique partition problem:

$$\min_{\mathtt{y}:\,\binom{S}{2}\to\{0,1\}}\sum_{abc\in\binom{S}{3}}\mathsf{c}_{abc}\,\mathtt{y}_{ab}\,\mathtt{y}_{bc}\,\mathtt{y}_{ac}+\sum_{ab\in\binom{S}{2}}\mathsf{c}_{ab}\,\mathtt{y}_{ab}+\mathsf{c}_{\emptyset}$$

subject to $y_{ab} + y_{bc} - 1 \le y_{ac}$ for all distinct $a, b, c \in S$.

Extended cost function c: $\binom{S}{3} \cup \binom{S}{2} \cup \emptyset \rightarrow \mathbb{R}$

Instance of the extended cubic clique partition problem:

$$\min_{\mathtt{y}:\,\binom{S}{2}\to\{0,1\}}\sum_{abc\in\binom{S}{3}}\mathsf{c}_{abc}\,\mathtt{y}_{ab}\,\mathtt{y}_{bc}\,\mathtt{y}_{ac}+\sum_{ab\in\binom{S}{2}}\mathsf{c}_{ab}\,\mathtt{y}_{ab}+\mathsf{c}_{\emptyset}$$

subject to $y_{ab} + y_{bc} - 1 \le y_{ac}$ for all distinct $a, b, c \in S$.

Construct **Improving Maps** for the clustering y

 \rightarrow Partial Optimality Conditions:

Extended cost function c: $\binom{S}{3} \cup \binom{S}{2} \cup \emptyset \rightarrow \mathbb{R}$

Instance of the extended cubic clique partition problem:

$$\min_{\mathtt{y}:\,\binom{S}{2}\to\{0,1\}}\sum_{abc\in\binom{S}{3}}\mathsf{c}_{abc}\,\mathtt{y}_{ab}\,\mathtt{y}_{bc}\,\mathtt{y}_{ac}+\sum_{ab\in\binom{S}{2}}\mathsf{c}_{ab}\,\mathtt{y}_{ab}+\mathsf{c}_{\emptyset}$$

subject to $y_{ab} + y_{bc} - 1 \le y_{ac}$ for all distinct $a, b, c \in S$.

- → Partial Optimality Conditions:
 - Subproblem-CUT-condition (cut subset from its complement)

Extended cost function c: $\binom{S}{3} \cup \binom{S}{2} \cup \emptyset \rightarrow \mathbb{R}$

Instance of the extended cubic clique partition problem:

$$\min_{\mathtt{y}:\,\binom{S}{2}\to\{0,1\}}\sum_{abc\in\binom{S}{3}}\mathsf{c}_{abc}\,\mathtt{y}_{ab}\,\mathtt{y}_{bc}\,\mathtt{y}_{ac}+\sum_{ab\in\binom{S}{2}}\mathsf{c}_{ab}\,\mathtt{y}_{ab}+\mathsf{c}_{\emptyset}$$

subject to $y_{ab} + y_{bc} - 1 \le y_{ac}$ for all distinct $a, b, c \in S$.

- → Partial Optimality Conditions:
 - Subproblem-CUT-condition (cut subset from its complement)
 - CUT-conditions (cut pairs and triples)

Extended cost function c: $\binom{S}{3} \cup \binom{S}{2} \cup \emptyset \rightarrow \mathbb{R}$

Instance of the extended cubic clique partition problem:

$$\min_{\mathtt{y}:\,\binom{S}{2}\to\{0,1\}}\sum_{abc\in\binom{S}{3}}\mathsf{c}_{abc}\,\mathtt{y}_{ab}\,\mathtt{y}_{bc}\,\mathtt{y}_{ac}+\sum_{ab\in\binom{S}{2}}\mathsf{c}_{ab}\,\mathtt{y}_{ab}+\mathsf{c}_{\emptyset}$$

subject to $y_{ab} + y_{bc} - 1 \le y_{ac}$ for all distinct $a, b, c \in S$.

- → Partial Optimality Conditions:
 - Subproblem-CUT-condition (cut subset from its complement)
 - 2 CUT-conditions (cut pairs and triples)
 - JOIN-conditions (join subsets, pairs and triples)

Extended cost function c: $\binom{S}{3} \cup \binom{S}{2} \cup \emptyset \rightarrow \mathbb{R}$

Instance of the extended cubic clique partition problem:

$$\min_{\mathtt{y}:\,\binom{S}{2}\to\{0,1\}}\sum_{abc\in\binom{S}{3}}\mathsf{c}_{abc}\,\mathtt{y}_{ab}\,\mathtt{y}_{bc}\,\mathtt{y}_{ac}+\sum_{ab\in\binom{S}{2}}\mathsf{c}_{ab}\,\mathtt{y}_{ab}+\mathsf{c}_{\emptyset}$$

subject to $y_{ab} + y_{bc} - 1 \le y_{ac}$ for all distinct $a, b, c \in S$.

Construct **Improving Maps** for the clustering y

- → Partial Optimality Conditions:
 - Subproblem-CUT-condition (cut subset from its complement)
 - 2 CUT-conditions (cut pairs and triples)
 - JOIN-conditions (join subsets, pairs and triples)

CUT-conditions can be applied simultaneously JOIN-conditions cannot be applied simultaneously!

Extended cost function c: $\binom{S}{3} \cup \binom{S}{2} \cup \emptyset \rightarrow \mathbb{R}$

Instance of the extended cubic clique partition problem:

$$\min_{\mathtt{y} \colon \binom{S}{2} \to \{0,1\}} \sum_{abc \in \binom{S}{3}} \mathsf{c}_{abc} \, \mathtt{y}_{ab} \, \mathtt{y}_{bc} \, \mathtt{y}_{ac} + \sum_{ab \in \binom{S}{2}} \mathsf{c}_{ab} \, \mathtt{y}_{ab} + \mathsf{c}_{\emptyset}$$

subject to $y_{ab} + y_{bc} - 1 \le y_{ac}$ for all distinct $a, b, c \in S$.

Construct Improving Maps for the clustering y

- → Partial Optimality Conditions:
 - Subproblem-CUT-condition (cut subset from its complement)
 - CUT-conditions (cut pairs and triples)
 - JOIN-conditions (join subsets, pairs and triples)

CUT-conditions can be applied simultaneously JOIN-conditions cannot be applied simultaneously!

Apply partial optimality connditions \rightarrow solve subproblems

Partial Optimality Algorithm

```
Partial Optimality Algorithm:
```

Input: clustering y without fixed labels
while condition applied do
apply subproblem-CUT-condition exhaustively
apply one of JOIN-conditions (in effective order)

end while

apply CUT-conditions exhaustively

Output: partially optimal clustering y with some fixed labels

Partial Optimality Algorithm

Partial Optimality Algorithm:

Input: clustering y without fixed labels

while condition applied do

apply subproblem-CUT-condition exhaustively apply one of JOIN-conditions (in effective order)

end while

apply CUT-conditions exhaustively

Output: partially optimal clustering y with some fixed labels

Reduction to subproblems:

① Subproblem-CUT-condition: fix CUT labels for element pairs from different sample subsets; solve each subset as an independent problem and accumulate the results in c_{\emptyset} ;

Partial Optimality Algorithm

Partial Optimality Algorithm:

Input: clustering y without fixed labels

while condition applied do

apply subproblem-CUT-condition exhaustively apply one of JOIN-conditions (in effective order)

end while

apply CUT-conditions exhaustively

Output: partially optimal clustering y with some fixed labels

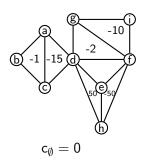
Reduction to subproblems:

- Subproblem-CUT-condition: fix CUT labels for element pairs from different sample subsets; solve each subset as an independent problem and accumulate the results in c_∅;
- ② JOIN-Conditions: fix JOIN labels for elements of the sample subset; add the join-cost to c_{\emptyset} ; solve the problem where the subset is considered as one sample;

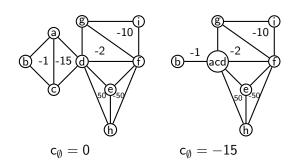
Subproblem-CUT: cut sample subsets R_1, R_2, \dots, R_k that are only connected via non-negative costs (applied if k > 1)

Subproblem-CUT: cut sample subsets R_1, R_2, \ldots, R_k that are only connected via non-negative costs (applied if k > 1)

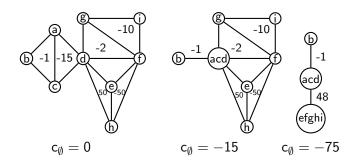
Subproblem-CUT: cut sample subsets R_1, R_2, \dots, R_k that are only connected via non-negative costs (applied if k > 1)



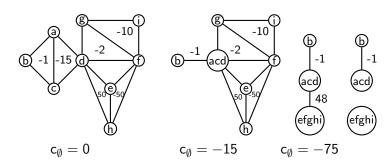
Subproblem-CUT: cut sample subsets R_1, R_2, \ldots, R_k that are only connected via non-negative costs (applied if k > 1)



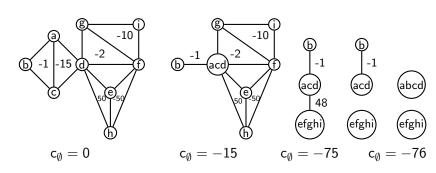
Subproblem-CUT: cut sample subsets R_1, R_2, \ldots, R_k that are only connected via non-negative costs (applied if k > 1)



Subproblem-CUT: cut sample subsets $R_1, R_2, ..., R_k$ that are only connected via non-negative costs (applied if k > 1)



Subproblem-CUT: cut sample subsets $R_1, R_2, ..., R_k$ that are only connected via non-negative costs (applied if k > 1)



Pair-JOIN-1: join samples i, j if their overall joining reward \geq the sum of rewards and penalties for joining some subset R with $i \in R$ and \overline{R} with $j \in \overline{R}$ (\approx i-j min-cut)

Pair-JOIN-1: join samples i, j if their overall joining reward \geq the sum of rewards and penalties for joining some subset R with $i \in R$ and \overline{R} with $j \in \overline{R}$ (\approx i-j min-cut)

Pair-JOIN-2: join samples i, k if there exists a sample triple ijk that fulfills 3 conditions (\approx i-jk min-cut, \approx ij-k min-cut, 1 explicit condition)

Pair-JOIN-1: join samples i, j if their overall joining reward \geq the sum of rewards and penalties for joining some subset R with $i \in R$ and \overline{R} with $j \in \overline{R}$ (\approx i-j min-cut)

Pair-JOIN-2: join samples i, k if there exists a sample triple ijk that fulfills 3 conditions (\approx i-jk min-cut, \approx ij-k min-cut, 1 explicit condition)

Pair-JOIN-3: join samples i, j if $c_{\{i,j\}} \le$ the sum of reward costs for joining pairs and triples containing i or j

Pair-JOIN-1: join samples i, j if their overall joining reward \geq the sum of rewards and penalties for joining some subset R with $i \in R$ and \overline{R} with $j \in \overline{R}$ (\approx i-j min-cut)

Pair-JOIN-2: join samples i, k if there exists a sample triple ijk that fulfills 3 conditions (\approx i-jk min-cut, \approx ij-k min-cut, 1 explicit condition)

Pair-JOIN-3: join samples i, j if $c_{\{i,j\}} \le$ the sum of reward costs for joining pairs and triples containing i or j

Pair-JOIN-4: join samples i, k if there exists a sample triple ijk such that 7 explicit conditions hold

Other JOIN-conditions

Pair-JOIN-1: join samples i, j if their overall joining reward \geq the sum of rewards and penalties for joining some subset R with $i \in R$ and \overline{R} with $j \in \overline{R}$ (\approx i-j min-cut)

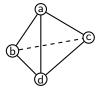
Pair-JOIN-2: join samples i, k if there exists a sample triple ijk that fulfills 3 conditions (\approx i-jk min-cut, \approx ij-k min-cut, 1 explicit condition)

Pair-JOIN-3: join samples i, j if $c_{\{i,j\}} \le$ the sum of reward costs for joining pairs and triples containing i or j

Pair-JOIN-4: join samples i, k if there exists a sample triple ijk such that 7 explicit conditions hold

Triple-JOIN: join samples i, j, k if the condition holds (similar to Pair-JOIN-1) (\approx i-jk min-cut)

$$\begin{aligned} c_{\{b,c,d\}} &= 10 \\ c_{\{a,b,c\}} &= c_{\{a,b,d\}} = c_{\{a,c,d\}} = -50 \end{aligned}$$

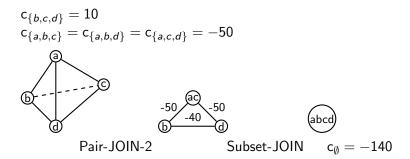


$$c_{\{b,c,d\}} = 10$$

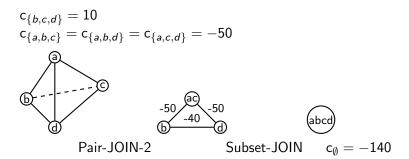
$$c_{\{a,b,c\}} = c_{\{a,b,d\}} = c_{\{a,c,d\}} = -50$$

Pair-JOIN-2

$$c_{\{b,c,d\}} = 10 \\ c_{\{a,b,c\}} = c_{\{a,b,d\}} = c_{\{a,c,d\}} = -50$$



Pair-CUT: cut samples i,j if the direct joing penalty \geq the sum of rewards for joining some subset R with $i \in R$ and \overline{R} with $j \in \overline{R}$ (\approx i-j min-cut)



Pair-CUT: cut samples i,j if the direct joing penalty \geq the sum of rewards for joining some subset R with $i \in R$ and \overline{R} with $j \in \overline{R}$ (\approx i-j min-cut)

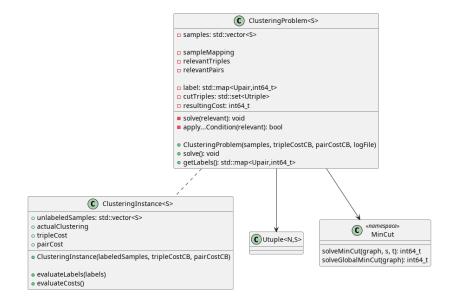
Triple-CUT: cut samples i, j, k if the condition holds (similar to Pair-CUT) (\approx i-jk min-cut)

Pair-CUT: cut samples i,j if the direct joing penalty \geq the sum of rewards for joining some subset R with $i \in R$ and \overline{R} with $j \in \overline{R}$ (\approx i-j min-cut)

Triple-CUT: cut samples i, j, k if the condition holds (similar to Pair-CUT) (\approx i-jk min-cut)

Samples in the pyramid with $c_{\{b,c,d\}}=100$ are unjoinable! Triple-CUT is applied to the triple bcd

Program Structure



Introduction

2 Partial Optimality for Cubic Clique Partition Problem

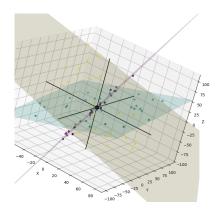
3 Cubic Subspace Instance Construction

4 Experiments and Evaluation

Conclusion

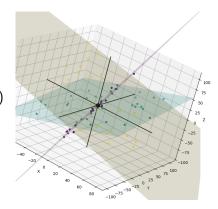
Plane Generation:

• generate 3 planes as distinct normal vectors $\vec{n}_1, \vec{n}_2, \vec{n}_3$ (normalized)



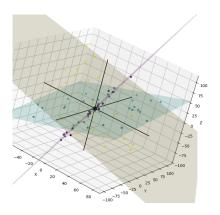
Plane Generation:

- generate 3 planes as distinct normal vectors $\vec{n}_1, \vec{n}_2, \vec{n}_3$ (normalized)
- compute a $\vec{r}_{i,1}$ (normalized) orthogonal to \vec{n}_i $(i \in \{1,2,3\})$



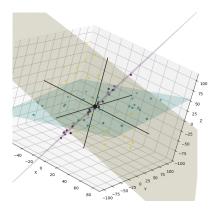
Plane Generation:

- generate 3 planes as distinct normal vectors $\vec{n}_1, \vec{n}_2, \vec{n}_3$ (normalized)
- compute a $\vec{r}_{i,1}$ (normalized) orthogonal to \vec{n}_i $(i \in \{1,2,3\})$
- compute the $\vec{r_{i,2}}$ (normalized) orthogonal to $\vec{n_i}$ and $\vec{r_{i,1}}$



Plane Generation:

- generate 3 planes as distinct normal vectors $\vec{n}_1, \vec{n}_2, \vec{n}_3$ (normalized)
- compute a $\vec{r}_{i,1}$ (normalized) orthogonal to \vec{n}_i $(i \in \{1,2,3\})$
- compute the $\vec{r}_{i,2}$ (normalized) orthogonal to \vec{n}_i and $\vec{r}_{i,1}$



Point Generation on the plane $(\vec{n}, \vec{r_1}, \vec{r_2})$, parameters (D, σ) :

- random variables $k_1, k_2 \in [-D, D]$ (uniform distribution)
- random variable k_n (normal distribution based on σ)
- generate point $p = k_1 \vec{r_1} + k_2 \vec{r_2} + k_n \vec{n}$

Cost Function

Triangle $abc \in \binom{S}{3}$

- **1** Smallest side $s < D/2 \rightarrow c_{abc} = 0$
- 2 Largest angle $\alpha > 150^{\circ} \rightarrow c_{abc} = 0$
- **③** ha, hb, hc: distances to the best fitting plane $ha + hb + hc > 3\sigma + 10^{-6}$ $→ c_{abc} = \frac{(ha+hb+hc)-(3\sigma+10^{-6})}{3D}$
- ho: distance from the origin to the triangle plane ho $> \frac{10\sigma}{\#points} + 10^{-6} \rightarrow c_{abc} = 0$
- of for all points p: hp: distance to the best fitting plane choose p if $hp < \sigma + 10^{-6}$ and $|\vec{p}| > 0.3D$ hp': distance to the best fitting plane of all chosen points $\delta_p = \frac{hp' (\sigma + 10^{-6})}{D}$, $SAME = \{p \colon \delta_p < 0\}$, $rew = \sum_{p \in SAME} \delta_p$, $|SAME| \le 3 \to c_{abc} = 0$ else $\to c_{abc} = 2^{|SAME| 4} rew$

Introduction

2 Partial Optimality for Cubic Clique Partition Problem

3 Cubic Subspace Instance Construction

4 Experiments and Evaluation

6 Conclusion

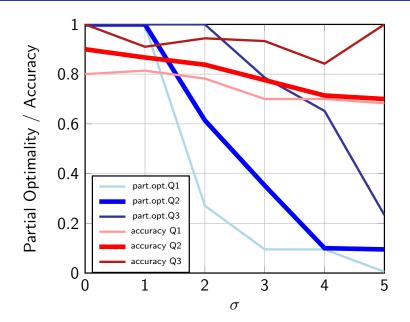
Experiments

My laptop characteristics, Random cubic subspace instances with different seeds max point component size 100 (noise are the percents then) no noise 0, small noise 1, significant noise 3, large noise 5, (Table: instance size + noise + instance count)

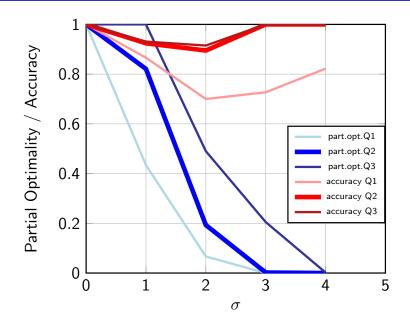
Cost Function Evaluation

blue and red dots, conflicts and and their effect (picture of the typical cost function evaluation)

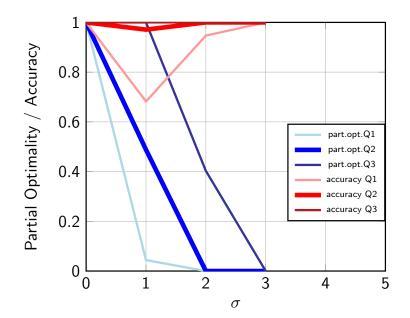
3x7 Partial Optimality / Accuracy



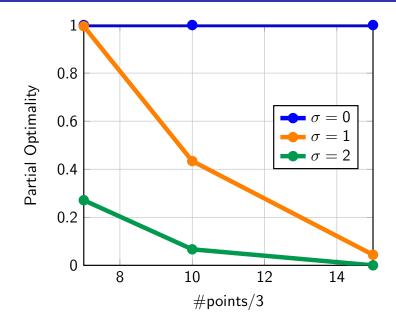
3x10 Partial Optimality / Accuracy



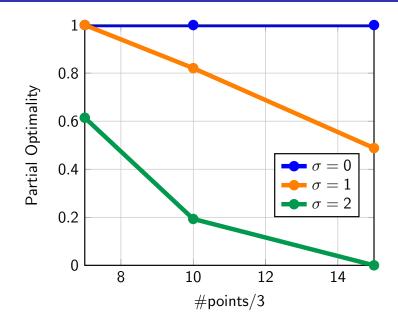
3x15 Partial Optimality / Accuracy



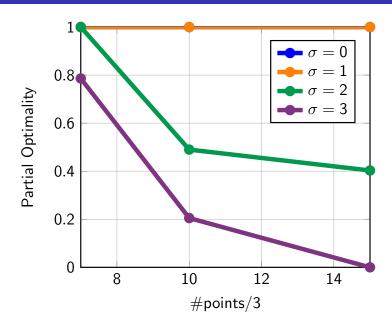
Partial Optimality (Q1)



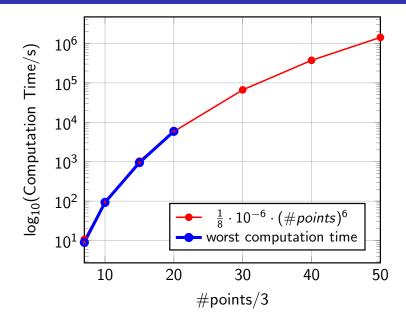
Partial Optimality (Q2)



Partial Optimality (Q3)



Computation Time (worst case)



Introduction

2 Partial Optimality for Cubic Clique Partition Problem

3 Cubic Subspace Instance Construction

4 Experiments and Evaluation

6 Conclusion

Conclusion

TODO: summarize the research results Partial optimality reimplementation, Dedicated Cost Function, Efficiency and Accuracy (shown by the experiments)
Future Work:

References

TODO: summarize the research results Partial optimality reimplementation, Dedicated Cost Function, Efficiency and Accuracy (shown by the experiments)
Future Work: