# 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

Research Results

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

$$\min_{\mathsf{y}:\;\binom{S}{2}\to\{0,1\}}\sum_{\{a,b,c\}\in\binom{S}{3}}\mathsf{c}_{\{a,b,c\}}\,\mathsf{y}_{\{a,b\}}\,\mathsf{y}_{\{b,c\}}\,\mathsf{y}_{\{a,c\}}$$

subject to  $\mathbf{y}_{\{a,b\}} + \mathbf{y}_{\{b,c\}} - 1 \leq \mathbf{y}_{\{a,c\}}$  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**:

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

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

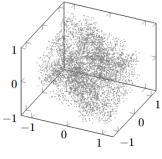
Find a **partially optimal solution**, i.e. fix some labels  $y_{\{a,b\}}$  for distinct  $a,b\in S$ 

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

in such way that there still exists an optimal solution.

**Subspace Instances** of the Cubic Clique Partition Problem

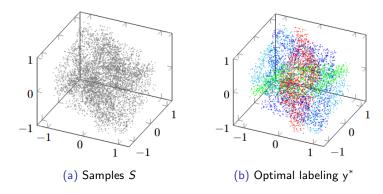
Samples S: points  $S \subset \mathbb{R}^3$ 



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  $\sigma$ 

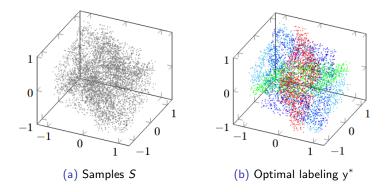


Subspace Instances of the Cubic Clique Partition Problem

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

Point generation: 3 distinct planes containing the origin, noise  $\sigma$ 

Optimal labeling y\*: original planes



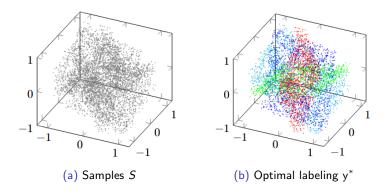
#### Subspace Instances of the Cubic Clique Partition Problem

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

Point generation: 3 distinct planes containing the origin, noise  $\sigma$ 

Optimal labeling y\*: original planes

Cost function c? (no concrete plane information given)



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

Task	Solution
read 'Partial Optimality in Cubic	my own implementation in
Correlation Clustering' [?]; imple-	$C{++}$ of the suggested al-
ment the algorithm for establish-	gorithm with some adjust-
ing partial optimality to the cubic	ments
clique partition problem;	

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 <b>suitable cost function</b> for the point triples	

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 <b>suitable cost function</b> for the point triples	point generation using the linear algebra methods; ex- perimentally determined ge- ometric cost function with a significant noise tolerance;

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 <b>suitable cost function</b> for the point triples	point generation using the linear algebra methods; ex- perimentally determined ge- ometric 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	

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 <b>suitable cost function</b> for the point triples	point generation using the linear algebra methods; ex- perimentally determined ge- ometric cost function with a significant noise tolerance;
apply the algorithm to the con- structed subspace instances, as- sess partial optimality, accuracy with respect to truth, and com- putation time	systematic empirical assess- ment of the algorithm appli- cation to the problem sub- space instances with my cost function proving its quality

Introduction

Partial Optimality for Cubic Clique Partition Problem

3 Cubic Subspace Instance Construction

4 Experiments and Evaluation

6 Research Results

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:

$$\min_{\mathtt{y} \colon \binom{S}{2} \to \{0,1\}} \sum_{\{a,b,c\} \in \binom{S}{3}} \mathsf{c}_{\{a,b,c\}} \, \mathsf{y}_{\{a,b\}} \, \mathsf{y}_{\{b,c\}} \, \mathsf{y}_{\{a,c\}} + \mathsf{c}_{\{a,b\}} \, \mathsf{y}_{\{a,b\}} + \mathsf{c}_{\emptyset}$$

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

Instance of the extended cubic clique partition problem:

$$\min_{\mathtt{y}:\;\binom{S}{2}\to\{0,1\}} \sum_{\{a,b,c\}\in\binom{S}{3}} \mathsf{c}_{\{a,b,c\}}\, \mathsf{y}_{\{a,b\}}\, \mathsf{y}_{\{b,c\}}\, \mathsf{y}_{\{a,c\}} + \mathsf{c}_{\{a,b\}}\, \mathsf{y}_{\{a,b\}} + \mathsf{c}_{\emptyset}$$

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

Instance of the extended cubic clique partition problem:

$$\min_{\mathtt{y} \colon \binom{S}{2} \to \{0,1\}} \sum_{\{a,b,c\} \in \binom{S}{3}} \mathsf{c}_{\{a,b,c\}} \, \mathsf{y}_{\{a,b\}} \, \mathsf{y}_{\{b,c\}} \, \mathsf{y}_{\{a,c\}} + \mathsf{c}_{\{a,b\}} \, \mathsf{y}_{\{a,b\}} + \mathsf{c}_{\emptyset}$$

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:

$$\min_{\mathbf{y}:\;\binom{S}{2}\to\{0,1\}} \sum_{\{a,b,c\}\in\binom{S}{3}} \mathbf{c}_{\{a,b,c\}}\,\mathbf{y}_{\{a,b\}}\,\mathbf{y}_{\{b,c\}}\,\mathbf{y}_{\{a,c\}} + \mathbf{c}_{\{a,b\}}\,\mathbf{y}_{\{a,b\}} + \mathbf{c}_{\emptyset}$$

- $\rightarrow$  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:

$$\min_{\mathbf{y}:\;\binom{S}{2}\to\{0,1\}} \sum_{\{a,b,c\}\in\binom{S}{3}} \mathbf{c}_{\{a,b,c\}}\,\mathbf{y}_{\{a,b\}}\,\mathbf{y}_{\{b,c\}}\,\mathbf{y}_{\{a,c\}} + \mathbf{c}_{\{a,b\}}\,\mathbf{y}_{\{a,b\}} + \mathbf{c}_{\emptyset}$$

- $\rightarrow$  Partial Optimality Conditions:
  - Subproblem-CUT-condition (cut subset from its complement)
  - 2 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:

$$\min_{\mathbf{y}: \; \binom{S}{2} \to \{0,1\}} \sum_{\{a,b,c\} \in \binom{S}{3}} \mathbf{c}_{\{a,b,c\}} \, \mathbf{y}_{\{a,b\}} \, \mathbf{y}_{\{b,c\}} \, \mathbf{y}_{\{a,c\}} + \mathbf{c}_{\{a,b\}} \, \mathbf{y}_{\{a,b\}} + \mathbf{c}_{\emptyset}$$

- $\rightarrow$  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:

$$\min_{\mathbf{y}:\;\binom{S}{2}\to\{0,1\}} \sum_{\{a,b,c\}\in\binom{S}{3}} \mathbf{c}_{\{a,b,c\}}\,\mathbf{y}_{\{a,b\}}\,\mathbf{y}_{\{b,c\}}\,\mathbf{y}_{\{a,c\}} + \mathbf{c}_{\{a,b\}}\,\mathbf{y}_{\{a,b\}} + \mathbf{c}_{\emptyset}$$

Construct **Improving Maps** for the clustering y

- $\rightarrow$  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:

$$\min_{\mathtt{y}: \, \binom{S}{2} \to \{0,1\}} \sum_{\{a,b,c\} \in \binom{S}{3}} \mathsf{c}_{\{a,b,c\}} \, \mathsf{y}_{\{a,b\}} \, \mathsf{y}_{\{b,c\}} \, \mathsf{y}_{\{a,c\}} + \mathsf{c}_{\{a,b\}} \, \mathsf{y}_{\{a,b\}} + \mathsf{c}_{\emptyset}$$

Construct Improving Maps for the clustering y

- $\rightarrow$  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!

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:

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

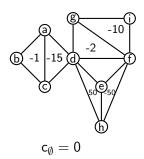
#### Reduction to subproblems:

- 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<sub>∅</sub>;
- 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;

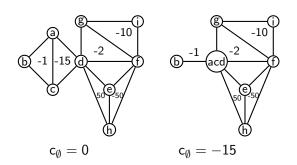
**Subproblem-CUT:** cut sample subsets  $R_1, R_2, \ldots, R_k$  that are only connected via non-negative costs (applied if k > 1)

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

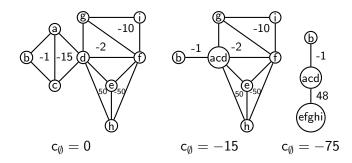
**Subproblem-CUT:** cut sample subsets  $R_1, R_2, \ldots, R_k$  that are only connected via non-negative costs (applied if k > 1)



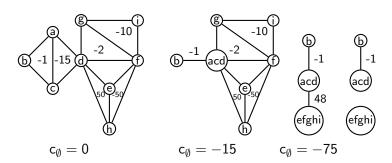
**Subproblem-CUT:** cut sample subsets  $R_1, R_2, ..., R_k$  that are only connected via non-negative costs (applied if k > 1)



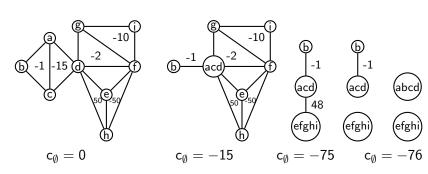
**Subproblem-CUT:** cut sample subsets  $R_1, R_2, ..., R_k$  that are only connected via non-negative costs (applied if k > 1)



**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  $R_1, R_2, \dots, R_k$  that are only connected via non-negative costs (applied if k > 1)



#### 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  $j \in \overline{R}$  (rhs  $\approx$  s-t min-cut)

#### 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  $j \in \overline{R}$  (rhs  $\approx$  s-t min-cut)

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

### 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  $j \in \overline{R}$  (rhs  $\approx$  s-t min-cut)

**Pair-JOIN-2:** join samples i, k if there exists a sample triple ijk that fulfills 3 conditions (2 conditions  $\approx$  s-t 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

### 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  $j \in \overline{R}$  (rhs  $\approx$  s-t min-cut)

**Pair-JOIN-2:** join samples i, k if there exists a sample triple ijk that fulfills 3 conditions (2 conditions  $\approx$  s-t 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  $j \in \overline{R}$  (rhs  $\approx$  s-t min-cut)

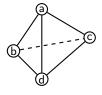
**Pair-JOIN-2:** join samples i, k if there exists a sample triple ijk that fulfills 3 conditions (2 conditions  $\approx$  s-t min-cut, 1 explicit condition)

**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) (rhs  $\approx$  s-t min-cut)

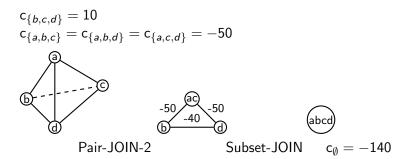
$$\begin{aligned} c_{\{b,c,d\}} &= 10 \\ c_{\{a,b,c\}} &= c_{\{a,b,d\}} = c_{\{a,c,d\}} = -50 \end{aligned}$$



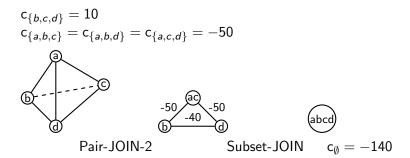
$$c_{\{b,c,d\}} = 10$$

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

$$c_{\{b,c,d\}} = 10 \\ c_{\{a,b,c\}} = c_{\{a,b,d\}} = c_{\{a,c,d\}} = -50$$
 
$$\frac{a}{b} - \frac{-50}{40} - \frac{a}{d} - \frac{-50}{40} - \frac{a}{d} - \frac{-50}{40} - \frac{-50}{40} - \frac{-140}{40} - \frac{-14$$



**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  $j \in \overline{R}$  (rhs  $\approx$  s-t 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  $j \in \overline{R}$  (rhs  $\approx$  s-t min-cut)

**Triple-CUT:** cutsamples i, j, k if the condition holds (similar to Pair-CUT) (rhs  $\approx$  s-t min-cut)

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

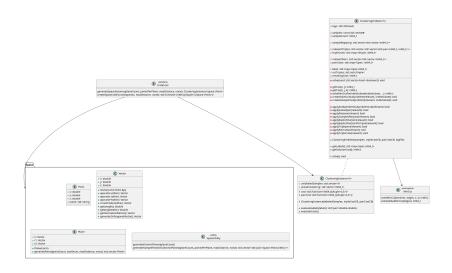
Pair-JOIN-2 Subset-JOIN  $c_{\emptyset} = -140$ 

**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  $j \in \overline{R}$  (rhs  $\approx$  s-t min-cut)

**Triple-CUT:** cutsamples i, j, k if the condition holds (similar to Pair-CUT) (rhs  $\approx$  s-t min-cut)

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

## Program Structure



# Pyramid Instance: Program Logs

```
Clustering problem has been inited with 4 relevant triples.
(negative: 3, positive: 1)
Trying independent subproblem cut (3.1)...
Trying subset join (3.11)...
Trying pair join (3.4)...
Trying complex pair join (3.6)...
* Applying the complex pair join (3.6)
Join: a c
Trying independent subproblem cut (3.1)...
Trying subset join (3.11)...
* Applying the subset join (3.11)
Join: ac b d
(0: cut; 1: joint; x: unknown)
                                         Unjoinable cluster triples:
Labeling:
                                          Clustering:
a b c d
a - 111
                                         Cluster 0: abcd
b 1 - 11
                                         Problem solved: completely
c 1 1 - 1
                                         Cost: -140
d 1 1 1 -
```

Introduction

2 Partial Optimality for Cubic Clique Partition Problem

3 Cubic Subspace Instance Construction

4 Experiments and Evaluation

Research Results

### Plane and Point Generation

plane Generation point Generation save as the labeled data, use the unlabeled data + cost function to init the ClusteringProblem

### Cost Function

My contribution!!! Cost function does not depend on the point distribution bounds!!! Steps to assign the cost !!!!!

Introduction

2 Partial Optimality for Cubic Clique Partition Problem

3 Cubic Subspace Instance Construction

4 Experiments and Evaluation

6 Research Results

### 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)

# Experiment Results for 3x7 Points

3x7 points: all results + time-optimality-accuracy (min-max-average)

## **Experiment Results**

3x(7x12x17x22) time-optimality-accuracy (min-max-average) DIAGRAM!!! mention the coefficients to prove the efficiency, partial optimality and accuracy!!!

Introduction

2 Partial Optimality for Cubic Clique Partition Problem

3 Cubic Subspace Instance Construction

4 Experiments and Evaluation

Research Results

### Research Results

Partial optimality reimplementation, Dedicated Cost Function, Efficiency and Accuracy (shown by the experiments)

### Future Work

Optimize the partial optimality algorithm Determine better parameters for the cost function Extend the cost function with more advanced conditions