The Two-Stage Approach for Solving Fair Clustering Problems

How to solve fair clustering problems?

- We are looking for algorithms with ***theoretical guarantees***:

  1-**Clustering Objective**:

  2-**The Fairness Constraint**:

How to solve fair clustering problems?

➢ We are looking for algorithms with ***theoretical guarantees***:

1-**Clustering Objective**:

$D = \min_{S,\varphi} \sum_{j \in \mathcal{C}} d^2(j, \varphi(j))$ → $\hat{D} \leq \alpha D$ (α>1 , recall NP-hardness)

2-**The Fairness Constraint**:

$$\begin{aligned} l_{blue}|C_i| &\leq \left|C_i^{blue}\right| \leq u_{blue}|C_i| \\ l_{red}|C_i| &\leq \left|C_i^{red}\right| \leq u_{red}|C_i| \end{aligned} \rightarrow \begin{aligned} (l_{blue}|C_i|) - \Delta &\leq \left|C_i^{blue}\right| \leq (u_{blue}|C_i|) + \Delta \\ (l_{red}|C_i|) - \Delta &\leq \left|C_i^{red}\right| \leq (u_{red}|C_i|) + \Delta \end{aligned}$$

-relax by $\Delta > 0$

How to solve fair clustering problems?

- We are looking for algorithms with ***theoretical guarantees*** over:

  1-**Clustering Objective**: $D = \min_{S,\varphi} \sum_{j \in \mathcal{C}} d^2(j, \varphi(j)) \rightarrow \widehat{D} \leq \alpha D$ (α>1 , recall NP-hardness)

  2-**Fairness Constraint**: $l_{blue}|C_i| \leq |C_i^{blue}| \leq u_{blue}|C_i|$ $\rightarrow$ $(l_{blue}|C_i|) - \Delta \leq |C_i^{blue}| \leq (u_{blue}|C_i|) + \Delta$

  $l_{red}|C_i| \leq |C_i^{red}| \leq u_{red}|C_i|$ $\quad$ $(l_{red}|C_i|) - \Delta \leq |C_i^{red}| \leq (u_{red}|C_i|) + \Delta$

- There is **NOT** a single approach to solve all fair variants.

  Unsurprising: Fair Clustering $\subset$ Constrained Clustering,
  No generic approach to solve Constrained Clustering for different constraints.

- Even the same problem maybe solved using different algorithms, e.g. Algorithm $\mathcal{A}_1$ has higher clustering quality than $\mathcal{A}_2$, but $\mathcal{A}_2$ has faster run time.

- For the k-(center,median, means): A simple approach with many applications → The two-stage approach.

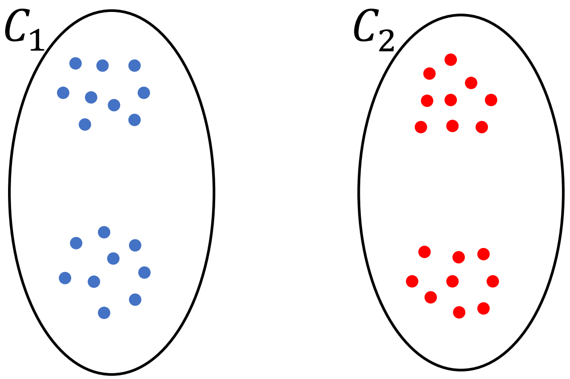
Two-Stage Approach

- ➢ **Step 1 (Open Centers)**: *Use a fairness-agnostic clustering algorithm* → this gives a collection of centers S

- ➢ **Step 2 (*Post-processing*)**: *process the clustering to satisfy the fairness constraint* at a bounded increase to the clustering cost (often that means carefully routing the points to the centers mostly using LP methods).

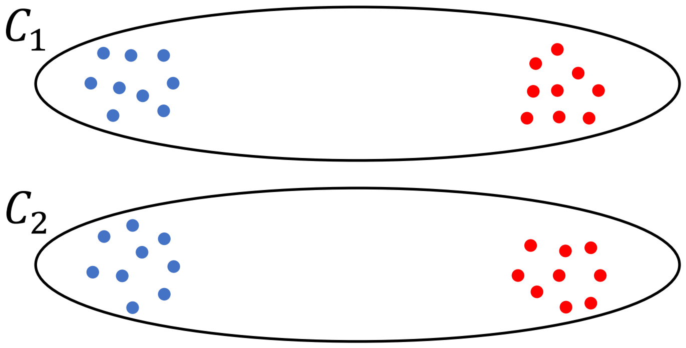
Two-Stage Approach: Group Fairness Example

- **Recall Group (demographic) Fairness**



Agnostic Clustering

$$\min \sum_{i=1}^{k} \sum_{j \in C_i} d(j, \mu_i)$$



Group Fair Clustering

$$\min \sum_{i=1}^{k} \sum_{j \in C_i} d(j, \mu_i)$$

s.t. $\quad l_{blue}|C_i| \leq |C_i^{blue}| \leq u_{blue}|C_i|$

$\quad l_{red}|C_i| \leq |C_i^{red}| \leq u_{red}|C_i|$

Two-Stage Approach: Group Fairness Example

Given Instance:

Two-Stage Approach: Group Fairness Example

Given Instance:

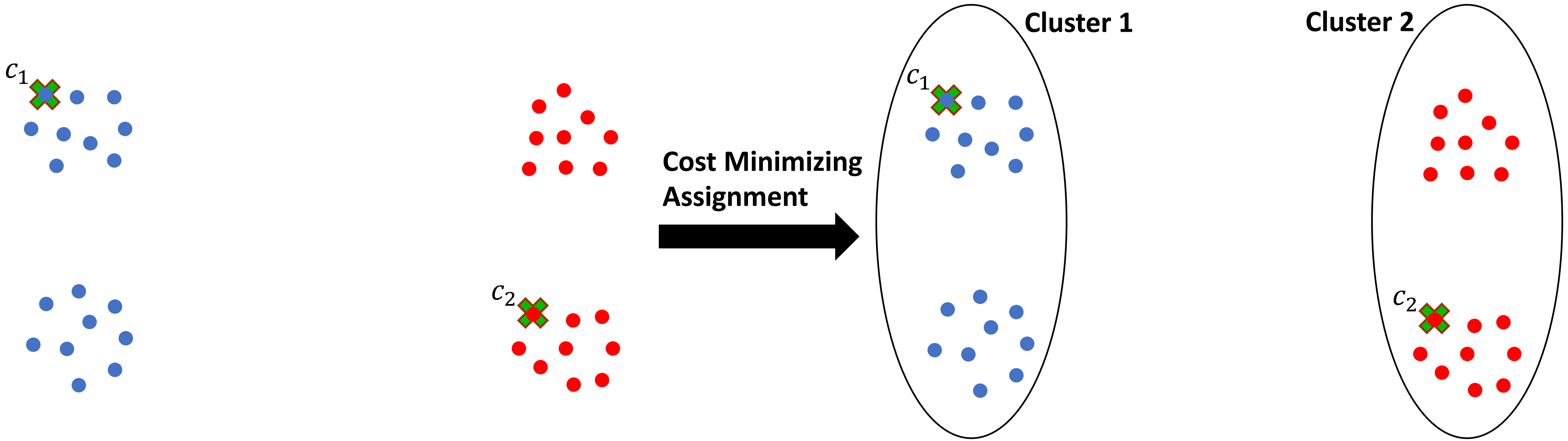
Two-Stage Approach: Group Fairness Example

- Centers are now **open!**
- How to assign points to centers??

Two-Stage Approach: Group Fairness Example

- Centers are now **open!**
- How to assign points to centers??

Cost minimizing assignment is <u>unfair</u> (clusters don't mix colors)



Two-Stage Approach: Group Fairness Example

- How to assign points to centers??

  (Step 2) Route points so as to **minimize clustering cost**

  subject to **satisfying color-proportional (fairness)** → **Setup an integer program**

Integer Program: $$\min_{x_{ij}} \sum_{i \in S} \sum_{j \in C} d(i,j) x_{ij}$$

$x_{ij} \in \{0,1\}$ 0-1 decision variable

$\sum_{i \in S} x_{ij} = x_{1j} + x_{2j} = 1$ point must be assigned to some center

$$l_{blue}\left(\sum_{j \in C} x_{1j}\right) \le \sum_{j \in C} p_j^{blue} x_{1j} \le u_{blue}\left(\sum_{j \in C} x_{1j}\right)$$

$$l_{red}\left(\sum_{j \in C} x_{1j}\right) \le \sum_{j \in C} p_j^{red} x_{1j} \le u_{red}\left(\sum_{j \in C} x_{1j}\right)$$

$$l_{blue}\left(\sum_{j \in C} x_{2j}\right) \le \sum_{j \in C} p_j^{blue} x_{2j} \le u_{blue}\left(\sum_{j \in C} x_{2j}\right)$$

$$l_{red}\left(\sum_{j \in C} x_{2j}\right) \le \sum_{j \in C} p_j^{red} x_{2j} \le u_{red}\left(\sum_{j \in C} x_{2j}\right)$$

Two-Stage Approach: Group Fairness Example

- How to assign points to centers??

  (Step 2) Route points so as to **minimize clustering cost**

  subject to **satisfying color-proportional (fairness)** → **Setup an integer program**

Integer Program: $$\min_{x_{ij}} \sum_{i \in S} \sum_{j \in C} d(i,j) x_{ij}$$

Integer Programs Generally Take Exponential Time!

$x_{ij} \in \{0,1\}$ 0-1 decision variable

$\sum_{i \in S} x_{ij} = x_{1j} + x_{2j} = 1$ point must be assigned to some center

$$l_{blue}\Big(\sum_{j \in C} x_{1j}\Big) \le \sum_{j \in C} p_j^{blue} x_{1j} \le u_{blue}\Big(\sum_{j \in C} x_{1j}\Big)$$

$$l_{red}\Big(\sum_{j \in C} x_{1j}\Big) \le \sum_{j \in C} p_j^{red} x_{1j} \le u_{red}\Big(\sum_{j \in C} x_{1j}\Big)$$

$$l_{blue}\Big(\sum_{j \in C} x_{2j}\Big) \le \sum_{j \in C} p_j^{blue} x_{2j} \le u_{blue}\Big(\sum_{j \in C} x_{2j}\Big)$$

$$l_{red}\Big(\sum_{j \in C} x_{2j}\Big) \le \sum_{j \in C} p_j^{red} x_{2j} \le u_{red}\Big(\sum_{j \in C} x_{2j}\Big)$$

Two-Stage Approach: Group Fairness Example

- How to assign points to centers??

(**Step 2)** Route points so as to **minimize clustering cost**

subject to **satisfying color-proportional (fairness)** → **Setup an integer program** → **Relax to LP**

Linear Program: $$\min_{x_{ij}} \sum_{i \in S} \sum_{j \in C} d(i,j) x_{ij}$$

$\cancel{x_{ij} \in \{0,1\}}$ $x_{ij} \in [0,1]$

$\sum_{i \in S} x_{ij} = x_{1j} + x_{2j} = 1$ point must be assigned to some center

$$l_{blue}\left(\sum_{j \in C} x_{1j}\right) \le \sum_{j \in C} p_j^{blue} x_{1j} \le u_{blue}\left(\sum_{j \in C} x_{1j}\right)$$

$$l_{red}\left(\sum_{j \in C} x_{1j}\right) \le \sum_{j \in C} p_j^{red} x_{1j} \le u_{red}\left(\sum_{j \in C} x_{1j}\right)$$

$$l_{blue}\left(\sum_{j \in C} x_{2j}\right) \le \sum_{j \in C} p_j^{blue} x_{2j} \le u_{blue}\left(\sum_{j \in C} x_{2j}\right)$$

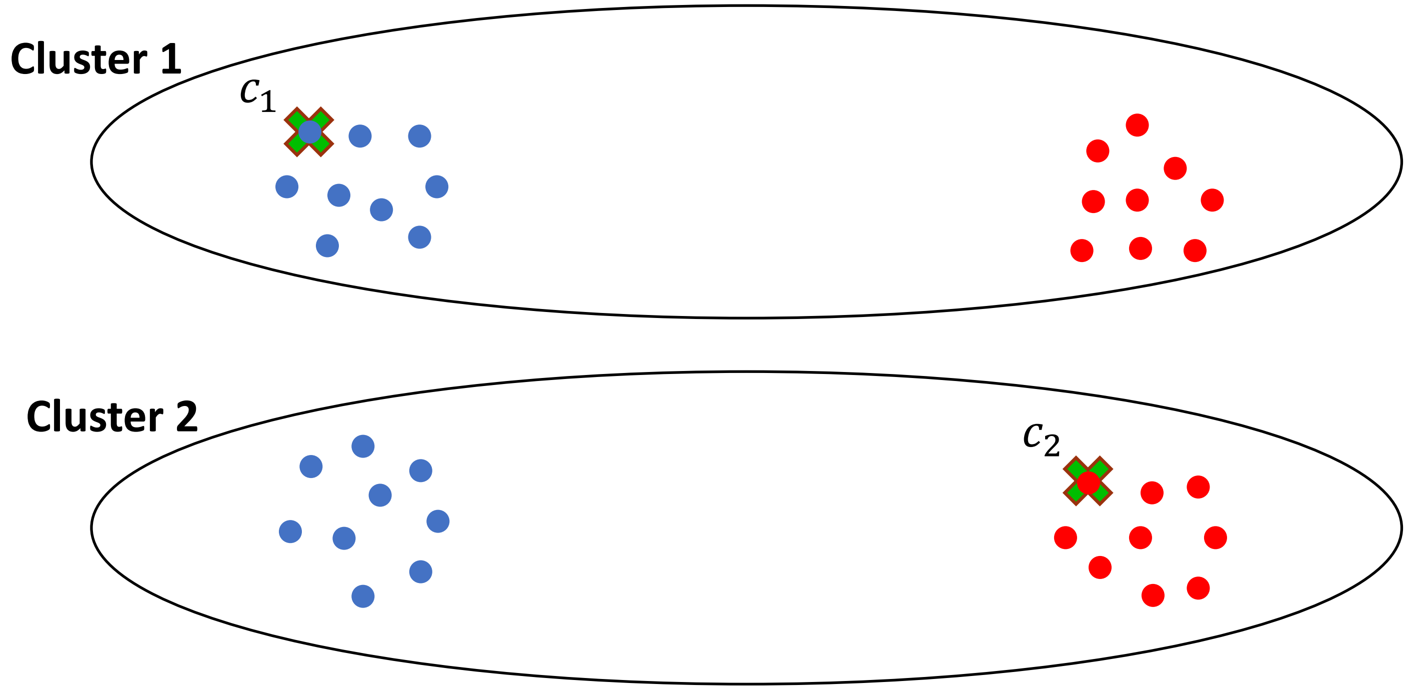
$$l_{red}\left(\sum_{j \in C} x_{2j}\right) \le \sum_{j \in C} p_j^{red} x_{2j} \le u_{red}\left(\sum_{j \in C} x_{2j}\right)$$

Two-Stage Approach: Group Fairness Example

- Resulting solution x_{ij} is possibly fractional (not 0 or 1)

Two-Stage Approach: Group Fairness Example

- Resulting solution x_{ij} is possibly fractional (not 0 or 1)
  - → Applying a rounding technique



Two-Stage Approach: Group Fairness Example

- Resulting solution x_{ij} is possibly fractional (not 0 or 1)

  → Applying a rounding technique

- Choice of rounding technique is non-trivial and often the most difficult step.

Two-Stage Approach

- Previous was for demographic fairness [Bera et al 2019; Bercea et al 2019; Esmaeili 2020].

- Other post processing approaches:

  -Combinatorial approach [Chakrabarti et al, AISTATS 2022]

  -Randomized approach [Brubach et al, ICML 2020]

THANK YOU!