Fair Clustering via Alignment

Kunwoong Kim, Jihu Lee, Sangchul Park, Yongdai Kim

ICML 2025 @ Vancouver, Canada

Paper

Code

Speaker: Kunwoong Kim

INDEX

1 Introduction & Contributions

2 Main results

3 Algorithm

4 Theoretical studies

5 Experiments

6 Conclusion

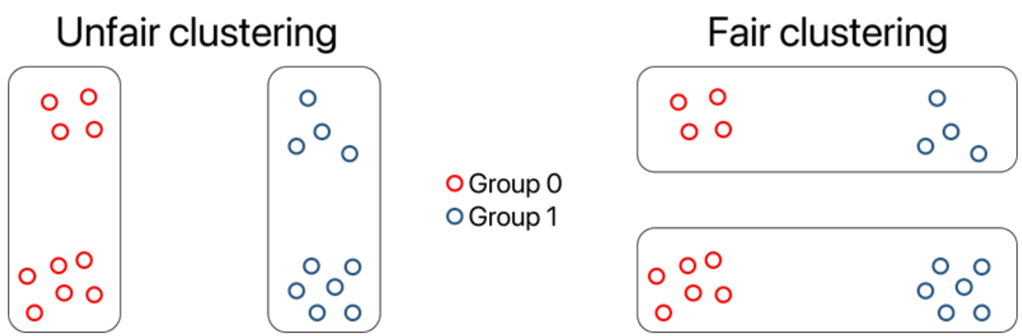
1

Introduction & Contributions

Fair Clustering

Group (Proportional) Fairness

Protected group ratio in each cluster ≈ Protected group ratio in the entire dataset



Why it matters?

Biased clustering —> Unfair downstream decisions

Examples: customer segmentation, medical cohorts

Existing works

Categories of fair clustering methods

Pre-processing

Build fair representation —> Apply clustering

In-processing

Jointly optimize clustering objective + fairness penalty

Post-processing

Find fair assignments given fixed cluster centers

Q. Can existing methods achieve the optimal trade-off between utility and fairness?

Contributions

- A novel decomposition of the fair K-means clustering cost:

Transport cost of building an aligned space

+ Clustering cost in that aligned space

- A new fair clustering algorithm (FCA), that is stable and guarantees convergence.

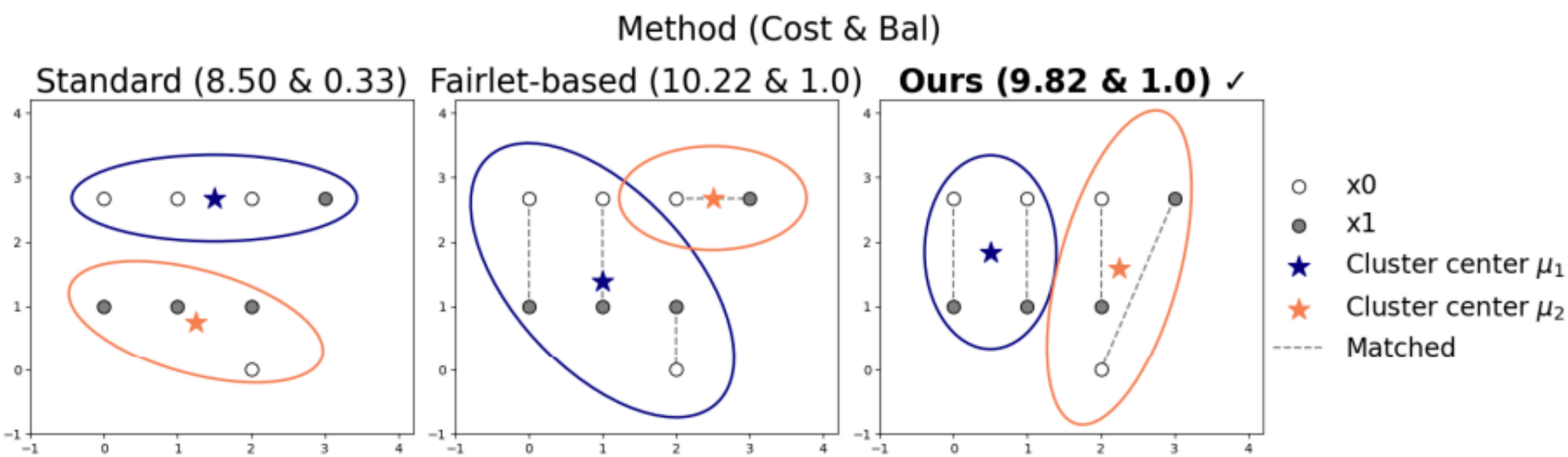
- Theoretically, FCA yields an approximately optimal fair clustering.
- Experimentally, FCA outperforms baseline fair clustering methods.

2

Main results

- **Idea**

  - Fair clustering can be found by matching.
  - How can we find matchings that yield the optimal fair clustering?



- ## Main results

If the sizes of the two protected groups are equal,
then there exists a one-to-one map between the two groups.

Theorem 3.1. *For any given perfectly fair deterministic assignment function* $\mathcal{A}$ *and cluster centers* $\boldsymbol{\mu}$*, there exists a one-to-one matching map* $\mathbf{T} : \mathcal{X}_s \to \mathcal{X}_{s'}$ *such that, for any* $s \in \{0, 1\}$, $C(\boldsymbol{\mu}, \mathcal{A}_0, \mathcal{A}_1) =$

$$\mathbb{E}_s \sum_{k=1}^{K} \mathcal{A}_s(\mathbf{X})_k \Bigg(\underbrace{\frac{\|\mathbf{X} - \mathbf{T}(\mathbf{X})\|^2}{4}}_{\text{Transport cost w.r.t. } \mathbf{T}} + \underbrace{\left\| \frac{\mathbf{X} + \mathbf{T}(\mathbf{X})}{2} - \mu_k \right\|^2}_{\text{Clustering cost w.r.t. } \boldsymbol{\mu} \text{ and } \mathbf{T}} \Bigg). \tag{2}$$

- ## Main results

Even when the sizes of the two protected groups are unequal,
we have a similar decomposition result using a stochastic matching map.

Let $\pi_s = n_s/(n_s + n_{s'})$ for $s \neq s' \in \{0, 1\}$. We then define

$$\mathbf{T}^{\mathrm{A}}(\mathbf{x}_0, \mathbf{x}_1) := \pi_0 \mathbf{x}_0 + \pi_1 \mathbf{x}_1$$

as the *alignment map*.

Theorem 3.3. *Let $\boldsymbol{\mu}^* \in \mathbb{R}^d$ and $\mathbb{Q}^* \in \mathcal{Q}$ be the cluster centers and joint distribution minimizing*

$$\mathbb{E}_{\mathbb{Q}} \left(2\pi_0\pi_1 \|\mathbf{X}_0 - \mathbf{X}_1\|^2 + \min_k \|\mathbf{T}^{\mathrm{A}}(\mathbf{X}_0, \mathbf{X}_1) - \mu_k\|^2 \right). \tag{3}$$

Then, $(\boldsymbol{\mu}^*, \mathcal{A}_0^*, \mathcal{A}_1^*)$ is the solution of the perfectly fair K-means clustering, where $\mathcal{A}_0^*(\mathbf{x})_k := \mathbb{Q}^* \left(\arg\min_{k'} \|\mathbf{T}^{\mathrm{A}}(\mathbf{x}, \mathbf{X}_1) - \mu_{k'}\|^2 = k | \mathbf{X}_0 = \mathbf{x} \right)$ and $\mathcal{A}_1^*(\mathbf{x})_k$ is defined similarly.

3

Algorithm

Overview

Optimal fair clustering can be found by simultaneously minimizing:

(i) The transport cost w.r.t. the matching between two groups (to align data points from two groups) and

(ii) The clustering cost w.r.t. the cluster centers in the aligned space.

Proposed algorithms

- FCA: perfect fairness
- FCA-C: control of fairness
- FCA-C is a general version of FCA.

Algorithm 1 FCA algorithm

input (i) Dataset $\mathcal{X}_0 \cup \mathcal{X}_1$. (ii) The number of clusters K.

1: Initialize cluster centers $\boldsymbol{\mu} = \{\mu_k\}_{k=1}^K$.

2: **while** $\boldsymbol{\mu}$ has not converged **do**

3: Update $\Gamma = [\gamma_{i,j}] \in \mathbb{R}_+^{n_0 \times n_1}$ by solving eq. (4) for a fixed $\boldsymbol{\mu}$. // Phase 1: update Γ

4: Update $\boldsymbol{\mu}$ by solving $\min_{\boldsymbol{\mu}} \sum_{i=1}^{n_0} \sum_{j=1}^{n_1} \gamma_{i,j} \min_k \|\mathbf{T}^{\mathrm{A}}(\mathbf{x}_i, \mathbf{x}_j) - \mu_k\|^2$ for a fixed Γ. // Phase 2: update $\boldsymbol{\mu}$

5: **end while**

6: Build fair assignments: for $\mathbf{x}_i \in \mathcal{X}_s$, define $\mathcal{A}_s(\mathbf{x}_i)_k := \sum_{\mathbf{x}_j \in \mathcal{X}_{s'}} n_s \gamma_{i,j} \mathbb{1}(\arg\min_{k'} \|\pi_s \mathbf{x}_i + \pi_{s'} \mathbf{x}_j - \mu_{k'}\|^2 = k), k \in [K]$.

output (i) Cluster centers $\boldsymbol{\mu} = \{\mu_k\}_{k=1}^K$. (ii) Assignments $\mathcal{A}_0(\mathbf{x}_i), \mathbf{x}_i \in \mathcal{X}_0$ and $\mathcal{A}_1(\mathbf{x}_j), \mathbf{x}_j \in \mathcal{X}_1$.

Algorithm 2 FCA-C algorithm

input (i) Dataset $\mathcal{X}_0 \cup \mathcal{X}_1$. (ii) The number of clusters K. (iii) Fairness level $\varepsilon \in [0, 1]$.

1: Initialize cluster centers $\boldsymbol{\mu} = \{\mu_k\}_{k=1}^K$ and a subset $\mathcal{W} \subset \mathcal{X}_0 \times \mathcal{X}_1$ such that $\frac{1}{n_0 n_1} \sum_{i=1}^{n_0} \sum_{j=1}^{n_1} \mathbb{I}((\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{W}) \leq \varepsilon$.

2: **while** $\boldsymbol{\mu}$ has not converged **do**

3: Calculate the costs $C_{K\text{-means}}$ for $(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{W}$ and C_{FCA} for $(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{W}^c$.

4: Update Γ by minimizing eq. (5) for fixed $\boldsymbol{\mu}$ and $\mathcal{W}$. // Phase 1: update Γ

5: Update $\boldsymbol{\mu}$ by minimizing eq. (5) for fixed Γ and $\mathcal{W}$. // Phase 2: update $\boldsymbol{\mu}$

6: For all $(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{X}_0 \times \mathcal{X}_1$, calculate $\eta(\mathbf{x}_i, \mathbf{x}_j) := 2\pi_0\pi_1\|\mathbf{x}_i - \mathbf{x}_j\|^2 + \min_k \|\mathbf{T}^{\mathrm{A}}(\mathbf{x}_i, \mathbf{x}_j) - \mu_k\|^2$. Let η_ε be the εth upper quantile. Update $\mathcal{W} = \{(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{X}_0 \times \mathcal{X}_1 : \eta(\mathbf{x}_i, \mathbf{x}_j) > \eta_\varepsilon\}$. // Phase 3: update $\mathcal{W}$

7: **end while**

8: Build fair assignment functions $\mathcal{A}_0$ and $\mathcal{A}_1$ following Equation (6).

output (i) Cluster centers $\boldsymbol{\mu} = \{\mu_k\}_{k=1}^K$. (ii) Assignments $\mathcal{A}_0(\mathbf{x}_i), \mathbf{x}_i \in \mathcal{X}_0$ and $\mathcal{A}_1(\mathbf{x}_j), \mathbf{x}_j \in \mathcal{X}_1$.

4

Theoretical studies

▪ Approximation guarantee

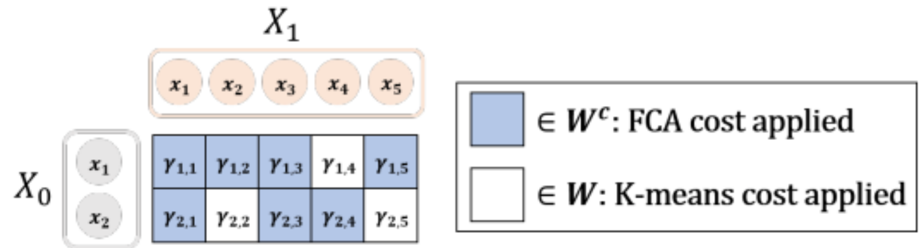
- FCA-C returns a $(\tau + 2)$-approximate solution, where τ is the approximation error of a standard clustering algorithm used to find initial cluster centers.

Suppose that $\sup_{\mathbf{x} \in \mathcal{X}} \|\mathbf{x}\|^2 \leq R$ for some $R > 0$.

Theorem 4.3 (Approximation guarantee of FCA-C)**.** *For any given* ε*, FCA-C algorithm returns an* $(\tau + 2)$*-approximate solution with a violation* $3R\varepsilon$ *for the optimal fair clustering, which is the solution of* $\min_{\boldsymbol{\mu}, \mathcal{A}_0, \mathcal{A}_1} C(\boldsymbol{\mu}, \mathcal{A}_0, \mathcal{A}_1)$ *subject to* $(\mathcal{A}_0, \mathcal{A}_1) \in \mathbf{A}_\varepsilon$.

- The rate $(\tau + 2)$ is similar to / better than existing algorithms.

- **Control of fairness level**



Theorem 4.1 (Equivalence between $\tilde{C}$ and constrained C). *Minimizing FCA-C objective $\tilde{C}(\mathbb{Q}, \mathcal{W}, \boldsymbol{\mu})$ with the corresponding assignment function defined in eq. (6), is equivalent to minimizing $C(\boldsymbol{\mu}, \mathcal{A}_0, \mathcal{A}_1)$ subject to $(\mathcal{A}_0, \mathcal{A}_1) \in \mathbf{A}_\varepsilon$.*

- **Balance bound**

Proposition 4.2 (Relationship between balance and ε). *For any assignment function $(\mathcal{A}_0, \mathcal{A}_1) \in \mathbf{A}_\varepsilon$, we have*

$$\max_{k \in [K]} \left| \frac{\sum_{\mathbf{x}_i \in \mathcal{X}_0} \mathcal{A}_0(\mathbf{x}_i)_k}{\sum_{\mathbf{x}_j \in \mathcal{X}_1} \mathcal{A}_1(\mathbf{x}_j)_k} - \frac{n_0}{n_1} \right| \leq c\varepsilon, \tag{7}$$

where $c = \frac{n_0}{n_1} \max_{k \in [K]} \frac{1}{\mathbb{E}_1 \mathcal{A}_1(\mathbf{X})_k}$.

- Balance is bounded by ϵ (i.e., the fairness level that FCA-C controls).

5

Experiments

Outperformance of FCA

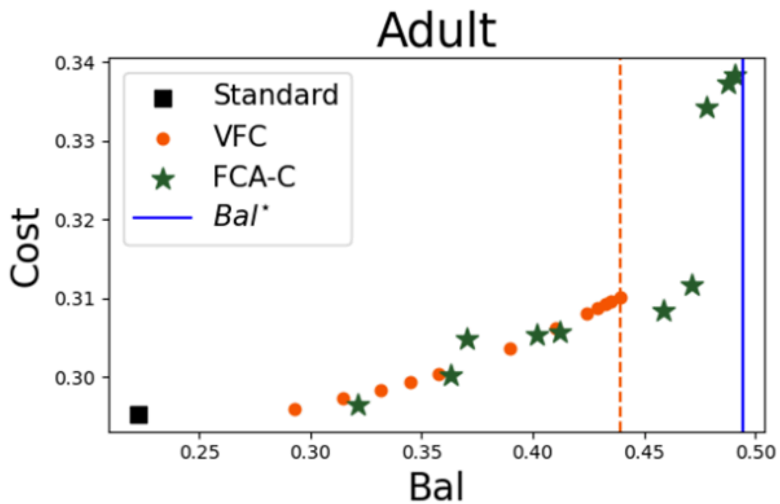
Tabular datasets

Dataset / Bal*	Adult / 0.494		Bank / 0.649		Census / 0.969	
With L_2 normalization	Cost (↓)	Bal (↑)	Cost (↓)	Bal (↑)	Cost (↓)	Bal (↑)
Standard (fair-unaware)	0.295	0.223	0.208	0.325	0.403	0.024
FCBC (Esmaeili et al., 2021)	0.314	0.443	0.685	0.615	1.006	0.926
SFC (Backurs et al., 2019)	0.534	0.489	0.410	0.632	1.015	0.937
FRAC (Gupta et al., 2023)	0.340	0.490	0.307	0.642	0.537	0.954
FCA ✓	**0.328**	0.493	**0.264**	0.645	**0.477**	0.962

Image datasets

Dataset / Bal*	RMNIST / 1.000			Office-31 / 0.282		
Performance	ACC (↑)	NMI (↑)	Bal (↑)	ACC (↑)	NMI (↑)	Bal (↑)
Standard (fair-unaware)	41.0	52.8	0.000	63.8	66.8	0.192
SFC (Backurs et al., 2019)	51.3	49.1	**1.000**	61.6	61.2	0.267
VFC (Ziko et al., 2021)	38.1	42.7	0.000	64.8	70.4	0.212
DFC (Li et al., 2020)	49.9	68.9	0.800	69.0	70.9	0.165
FCMI (Zeng et al., 2023)	88.4	**86.4**	0.995	**70.0**	**71.2**	0.226
FCA ✓	**89.0**	79.0	**1.000**	67.6	70.5	**0.270**

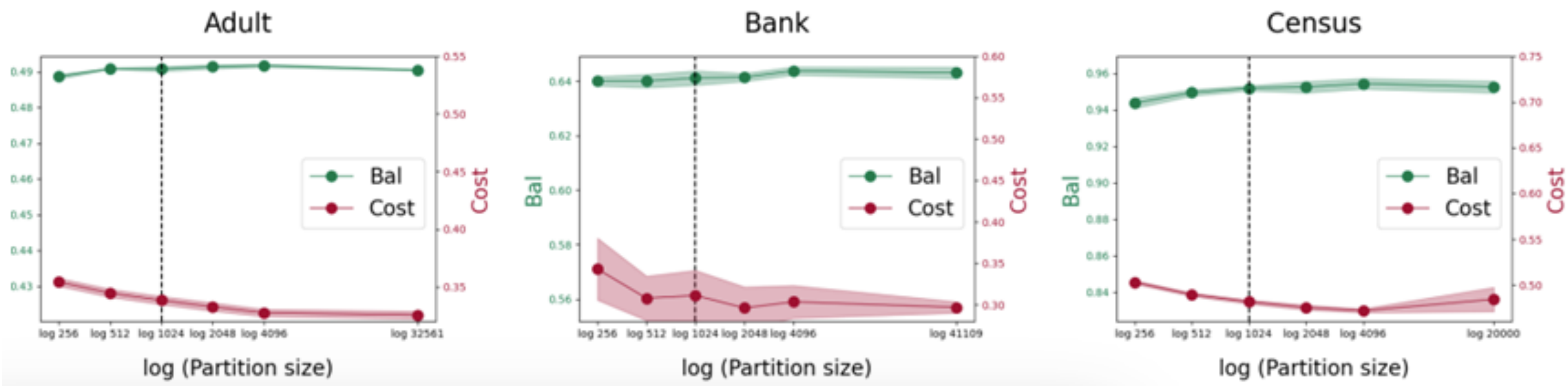
- **Fairness level control**



- **Stability / Robustness**

Dataset / Bal*	ADULT / 0.494		BANK / 0.649		CENSUS / 0.969	
With L_2 normalization	Cost	Bal	Cost	Bal	Cost	Bal
FCA (K-means++)	0.328	0.493	0.264	0.645	0.477	0.962
FCA (K-means random)	0.331	0.490	0.275	0.646	0.477	0.955
FCA (Gradient-based)	0.339	0.492	0.254	0.640	0.478	0.957

- ## Partitioning technique



- ## Linear program vs. Sinkhorn

ADULT			
Bal$^\star = 0.494$	Cost (↓)	Bal (↑)	Runtime / iteration (sec)
FCA (Sinkhorn, $\lambda = 1.0$)	0.350	0.271	4.98
FCA (Sinkhorn, $\lambda = 0.1$)	0.315	0.463	5.12
FCA (Sinkhorn, $\lambda = 0.01$)	0.330	0.491	5.55
FCA (Linear program)	0.328	0.493	5.67

6

Conclusion

Summary

- Decomposition: Alignment + Clustering
- FCA: stable and provable fair K-means clustering algorithm
- FCA-C: a variant of FCA, which can control fairness level

Future works

- Applying FCA to other clustering algorithms such as model-based clustering, e.g., Gaussian mixture.

Thank you!

Questions? Email me at:

kwkim.online@gmail.com