A Proof of Lemma 2

To prove Lemma 2 we start by proving a few inequalities. Since $\mathcal{A}$ is an $(\epsilon_1, \epsilon_2, \mathsf{Q})$-solver, using Definition 4 and Taylor's expansion, we get for any $i \in [n]$ and $j \in [k]$,

$$\widehat{\mathsf{U}}_{ij}^{\alpha} \leq (\mathsf{U}_{ij} + \epsilon_2)^{\alpha} \tag{11}$$

$$= \mathsf{U}_{ij}^{\alpha} + \alpha \mathsf{U}_{ij}^{\alpha-1}\epsilon_2 + o(\epsilon_2^2) \tag{12}$$

$$\leq \mathsf{U}_{ij}^{\alpha} + \alpha\epsilon_2 + o(\epsilon_2^2), \tag{13}$$

where the last inequality follows from the fact that $\mathsf{U}_{ij} \leq 1$. Also, for any $i \in [n]$ and $j \in [k]$, we have

$$\left\|\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}}_j\right\|_2^2 = \left\|\boldsymbol{x}_i - \boldsymbol{\mu}_j\right\|_2^2 + 2 \cdot (\boldsymbol{x}_i - \boldsymbol{\mu}_j)^T(\boldsymbol{\mu}_j - \widehat{\boldsymbol{\mu}}_j) + \left\|\boldsymbol{\mu}_j - \widehat{\boldsymbol{\mu}}_j\right\|_2^2 \tag{14}$$

$$\leq \left\|\boldsymbol{x}_i - \boldsymbol{\mu}_j\right\|_2^2 + 4 \cdot \mathsf{R} \cdot \epsilon_1 + \epsilon_1^2, \tag{15}$$

where the inequality follows from the fact that $\mathcal{A}$ is an $(\epsilon_1, \epsilon_2, \mathsf{Q})$-solver, the Cauchy-Schwarz inequality, and the fact that since $\boldsymbol{x}_i \in \mathcal{B}(0, \mathsf{R})$ so as $\mu_{\mathbf{j}} \in \mathcal{B}(\mathbf{0}, \mathsf{R})$, and thus $\left\|\boldsymbol{x}_i - \boldsymbol{\mu}_j\right\|_2^2 \leq 4 \cdot \mathsf{R}^2$. Finally, we have for any $i \neq j \in [k]$,

$$\left\|\widehat{\boldsymbol{\mu}}_i - \widehat{\boldsymbol{\mu}}_j\right\|_2^2 \geq \left\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right\|_2^2 + \left\|\widehat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_i\right\|_2^2 + \left\|\widehat{\boldsymbol{\mu}}_j - \boldsymbol{\mu}_j\right\|_2^2 + 2(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T(\widehat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_i)$$

$$+ 2(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T(\widehat{\boldsymbol{\mu}}_j - \boldsymbol{\mu}_j) + 2(\widehat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_i)^T(\widehat{\boldsymbol{\mu}}_j - \boldsymbol{\mu}_j) \tag{16}$$

$$\geq \left\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right\|_2^2 - 2\epsilon_1^2 - 8\mathsf{R}\epsilon_1 - 2\epsilon_1^2 \tag{17}$$

$$= \left\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right\|_2^2 - 8\mathsf{R}\epsilon_1 - 4\epsilon_1^2. \tag{18}$$

Now, with the above results, we note that

$$\mathsf{J}_{\mathsf{fm}}(\mathcal{X}, \widehat{\mathcal{P}}) = \sum_{i=1}^{n}\sum_{j=1}^{k} \widehat{\mathsf{U}}_{ij}^{\alpha} \left\|\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}}_j\right\|_2^2 \tag{19}$$

$$\leq \sum_{i=1}^{n}\sum_{j=1}^{k} \mathsf{U}_{ij}^{\alpha} \left\|\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}}_j\right\|_2^2 + \alpha\epsilon_2 \sum_{i=1}^{n}\sum_{j=1}^{k} \left\|\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}}_j\right\|_2^2 + o(\epsilon_2^2). \tag{20}$$

Now since $\boldsymbol{\mu}_j \in \mathcal{B}(0, \mathsf{R})$, we can say that $\widehat{\boldsymbol{\mu}}_j \in \mathcal{B}(0, \mathsf{R} + \epsilon_1)$. Therefore, $\left\|\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}}_j\right\|_2^2 \leq 2[\mathsf{R}^2 + (\mathsf{R} + \epsilon_1)^2]$. Hence,

$$\mathsf{J}_{\mathsf{fm}}(\mathcal{X}, \widehat{\mathcal{P}}) \leq \sum_{i=1}^{n}\sum_{j=1}^{k} \mathsf{U}_{ij}^{\alpha} \left\|\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}}_j\right\|_2^2 + 2nk\alpha\epsilon_2[\mathsf{R}^2 + (\mathsf{R} + \epsilon_1)^2] + o(\epsilon_2^2). \tag{21}$$

Next, using (15), we get

$$\sum_{i=1}^{n}\sum_{j=1}^{k} \mathsf{U}_{ij}^{\alpha} \left\|\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}}_j\right\|^2 \leq \sum_{i=1}^{n}\sum_{j=1}^{k} \mathsf{U}_{ij}^{\alpha} \left\|\boldsymbol{x}_i - \boldsymbol{\mu}_j\right\|_2^2 + \left[4 \cdot \mathsf{R} \cdot \epsilon_1 + \epsilon_1^2\right] \sum_{i=1}^{n}\sum_{j=1}^{k} \mathsf{U}_{ij}^{\alpha} \tag{22}$$

$$\leq \sum_{i=1}^{n}\sum_{j=1}^{k} \mathsf{U}_{ij}^{\alpha} \left\|\boldsymbol{x}_i - \boldsymbol{\mu}_j\right\|_2^2 + n\left[4 \cdot \mathsf{R} \cdot \epsilon_1 + \epsilon_1^2\right] \tag{23}$$

$$= \mathsf{J}_{\mathsf{fm}}(\mathcal{X}, \mathcal{P}) + n\left[4 \cdot \mathsf{R} \cdot \epsilon_1 + \epsilon_1^2\right], \tag{24}$$

where the second inequality follows from the fact that $\mathsf{U}_{ij} \in [0, 1]$, and thus $\sum_{i=1}^{n}\sum_{j=1}^{k} \mathsf{U}_{ij}^{\alpha} \leq \sum_{i=1}^{n}\sum_{j=1}^{k} \mathsf{U}_{ij} = n$. Therefore, we obtain

$$\mathsf{J}_{\mathsf{fm}}(\mathcal{X}, \widehat{\mathcal{P}}) \leq \mathsf{J}_{\mathsf{fm}}(\mathcal{X}, \mathcal{P}) + n\left[4 \cdot \mathsf{R} \cdot \epsilon_1 + \epsilon_1^2\right] + 2nk\alpha\epsilon_2[\mathsf{R}^2 + (\mathsf{R} + \epsilon_1)^2] + o(\epsilon_2^2) \tag{25}$$

$$\leq \mathsf{J}_{\mathsf{fm}}(\mathcal{X}, \mathcal{P}) + n \cdot O(\epsilon_1) + nk \cdot O(\epsilon_2) + n \cdot o(\epsilon_1^2) + nk \cdot o(\epsilon_2^2). \tag{26}$$

We are now in a position to bound $\mathsf{XB}(\mathcal{X}, \widehat{\mathcal{P}})$. Using (18) and (26), we have

$$\mathsf{XB}(\mathcal{X}, \widehat{\mathcal{P}}) = \frac{\mathsf{J}_{\mathsf{fm}}(\mathcal{X}, \widehat{\mathcal{P}})}{nk \cdot \min_{i\neq j} \left\|\widehat{\boldsymbol{\mu}}_i - \widehat{\boldsymbol{\mu}}_j\right\|_2^2} \tag{27}$$

$$\leq \frac{\mathsf{J}_{\mathsf{fm}}(\mathcal{X}, \mathcal{P}) + n \cdot O(\epsilon_1) + nk \cdot O(\epsilon_2) + n \cdot o(\epsilon_1^2) + nk \cdot o(\epsilon_2^2)}{nk \cdot \left[\min_{i\neq j} \left\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right\|_2^2 - 8\mathsf{R}\epsilon_1 - 4\epsilon_1^2\right]} \tag{28}$$

$$= \frac{\mathsf{J}_{\mathsf{fm}}(\mathcal{X}, \mathcal{P})}{nk \cdot \min_{i\neq j} \left\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right\|_2^2} + \frac{n \cdot O(\epsilon_1) + nk \cdot O(\epsilon_2) + n \cdot o(\epsilon_1^2) + nk \cdot o(\epsilon_2^2)}{nk \cdot \min_{i\neq j} \left\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right\|_2^2}$$

$$+ \frac{\mathsf{J}_{\mathsf{fm}}(\mathcal{X}, \mathcal{P})}{nk \cdot \min_{i\neq j} \left\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right\|_2^2} \frac{O(\epsilon_1)}{\min_{i\neq j} \left\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right\|_2^2} + o(\epsilon_1^2 + \epsilon_2^2) \tag{29}$$

$$= \mathsf{XB}(\mathcal{X}, \mathcal{P}) + \mathsf{XB}(\mathcal{X}, \mathcal{P}) \cdot \frac{O(\epsilon_1)}{\min_{i\neq j} \left\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right\|_2^2} + \frac{O(\epsilon_2)}{\min_{i\neq j} \left\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right\|_2^2}$$

$$+ o\left(\frac{\epsilon_1^2 + \epsilon_2^2}{\min_{i\neq j} \left\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right\|_2^2}\right). \tag{30}$$

Using the same steps a similar lower bound can be obtained, which concludes the proof.

B Auxiliary Lemmata

In this section we present and prove a few auxiliary results which will be used in the proofs our main results. We start with the following standard concentration inequalities.

Lemma 3 (Hoeffding's inequality). *Let $\mathsf{X}_1, \mathsf{X}_2, \ldots, \mathsf{X}_n$ be i.i.d random variables, such that $|\mathsf{X}_i| \leq \mathsf{R}$ a.s., and $\mathbb{E}\mathsf{X}_i = \mu$, for all $i \in [n]$. Then, with probability at least $1 - \delta$,*

$$\left|\frac{1}{n}\sum_{i=1}^{n} \mathsf{X}_i - \mu\right| \leq \mathsf{R}\epsilon, \tag{31}$$

if $n \geq \frac{c\log(1/\delta)}{2\epsilon^2}$, where $c > 0$ is some absolute constant.

Lemma 4 (Generalized Hoeffding's inequality). *Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be i.i.d random vectors, such that $\|\mathbf{X}_i\|_2 \leq \mathsf{R}$ a.s., and $\mathbb{E}\mathbf{X}_i = \boldsymbol{\mu}$, for all $i \in [n]$. Then, with probability at least $1 - \delta$,*

$$\left\|\frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i - \boldsymbol{\mu}\right\|_2^2 \leq \mathsf{R}^2\epsilon, \tag{32}$$

if $n \geq \frac{c\log(1/\delta)}{\epsilon^2}$, where $c > 0$ is some absolute constant.

The following locality lemma states that the fuzzy k-means function is strictly increasing.

Lemma 5. *Let $(\mathcal{X}, \mathcal{P}^\star)$ be a clustering instance, where $\mathcal{P}^\star$ refers to the optimal solution for the fuzzy k-mean problem (namely, minimizes the objective in (2)). Then, for any $i, j \in [n]$ and $\ell \in [k]$ with $\|\boldsymbol{x}_i - \boldsymbol{\mu}_\ell^\star\|_2^2 \leq \|\boldsymbol{x}_j - \boldsymbol{\mu}_\ell^\star\|_2^2$, we have $\mathsf{U}_{i\ell} \geq \mathsf{U}_{j\ell}$.*

Proof of Lemma 5. Consider some $i, j \in [n]$ and $\ell \in [k]$ with $\|\boldsymbol{x}_i - \boldsymbol{\mu}_\ell^\star\|_2^2 \leq \|\boldsymbol{x}_j - \boldsymbol{\mu}_\ell^\star\|_2^2$. By definition $\{\mathsf{U}_{i\ell}^\star\}_{i=1}^n$ minimizes the cost $\sum_{i=1}^n \mathsf{U}_{i\ell}^\alpha \|\boldsymbol{x}_i - \boldsymbol{\mu}_\ell\|_2^2$. This implies that,

$$\mathsf{U}_{i\ell}^\alpha \|\boldsymbol{x}_i - \boldsymbol{\mu}_\ell\|_2^2 + \mathsf{U}_{j\ell}^\alpha \|\boldsymbol{x}_j - \boldsymbol{\mu}_\ell\|_2^2 \leq \mathsf{U}_{j\ell}^\alpha \|\boldsymbol{x}_i - \boldsymbol{\mu}_\ell\|_2^2 + \mathsf{U}_{i\ell}^\alpha \|\boldsymbol{x}_j - \boldsymbol{\mu}_\ell\|_2^2, \tag{33}$$

which is equivalent to,

$$\mathsf{U}_{i\ell}^\alpha \left[\|\boldsymbol{x}_i - \boldsymbol{\mu}_\ell\|_2^2 - \|\boldsymbol{x}_j - \boldsymbol{\mu}_\ell\|_2^2\right] \leq \mathsf{U}_{j\ell}^\alpha \left[\|\boldsymbol{x}_i - \boldsymbol{\mu}_\ell\|_2^2 - \|\boldsymbol{x}_j - \boldsymbol{\mu}_\ell\|_2^2\right]. \tag{34}$$

Since $\|\boldsymbol{x}_i - \boldsymbol{\mu}_\ell\|_2^2 - \|\boldsymbol{x}_j - \boldsymbol{\mu}_\ell\|_2^2 \leq 0$, we get $\mathsf{U}_{i\ell}^\alpha \geq \mathsf{U}_{j\ell}^\alpha$, which concludes the proof. □

Algorithm 6 Algorithm for estimating the mean $\boldsymbol{\mu}_j$ for any $j \in [k]$.

Input: $\mathcal{X}$, $\mathcal{O}_{\mathsf{fuzzy}}$, α, and m.
Output: $\widehat{\boldsymbol{\mu}}_j$
1: Initialize $\mathcal{S} \leftarrow \phi$.
2: **for** $s = 1, 2, \ldots, m$ **do**
3: Sample i uniformly at random from $[n]$ and update $\mathcal{S} \leftarrow \mathcal{S} \cup \{i\}$.
4: Query $\mathcal{O}_{\mathsf{fuzzy}}(i, j)$.
5: **end for**
6: Compute $\widehat{\boldsymbol{\mu}}_j = \frac{\sum_{i\in\mathcal{S}} \mathsf{U}_{ij}^{\alpha} \boldsymbol{x}_i}{\sum_{i\in\mathcal{S}} \mathsf{U}_{ij}^{\alpha}}$.

Next, we analyze the performance of Algorithm 6, which estimates the center of a given cluster using a set of randomly sampled elements. Note that this algorithm is used as a sub-routine in Algorithm 1.

Lemma 6 (Estimate of mean using uniform sampling)**.** *Let $(\mathcal{X}, \mathcal{P})$ be a consistent center-based clustering instance, and let $\delta \in (0, 1)$. With probability at least $1 - \delta$, Algorithm 6 outputs an estimate $\widehat{\boldsymbol{\mu}}_j$ such that*

$$\left\|\boldsymbol{\mu}_j - \widehat{\boldsymbol{\mu}}_j\right\|_2^2 \leq \frac{4\mathsf{R}^2}{\mathsf{Y}^2}\sqrt{\frac{c}{m}\log\frac{1}{\delta}}, \tag{35}$$

where $\mathsf{Y} \triangleq \min_{j\in[k]} \frac{1}{n}\sum_{i\in[n]} \mathsf{U}_{ij}^{\alpha}$, and $c > 0$ is some absolute constant.

Proof of Lemma 6. First, note that

$$\widehat{\boldsymbol{\mu}}_j = \frac{\sum_{i\in\mathcal{S}} \mathsf{U}_{ij}^{\alpha}\boldsymbol{x}_i}{\sum_{i\in\mathcal{S}} \mathsf{U}_{ij}^{\alpha}} = \frac{(1/m)\sum_{i\in\mathcal{S}} \mathsf{U}_{ij}^{\alpha}\boldsymbol{x}_i}{(1/m)\sum_{i\in\mathcal{S}} \mathsf{U}_{ij}^{\alpha}} \triangleq \frac{\bar{\boldsymbol{\lambda}}_{\boldsymbol{x}}}{\bar{\mathsf{Y}}}. \tag{36}$$

Recall that the true mean of the j^{th} cluster is

$$\boldsymbol{\mu}_j = \frac{\sum_{i\in[n]} \mathsf{U}_{ij}^{\alpha}\boldsymbol{x}_i}{\sum_{i\in[n]} \mathsf{U}_{ij}^{\alpha}} = \frac{(1/n)\sum_{i\in[n]} \mathsf{U}_{ij}^{\alpha}\boldsymbol{x}_i}{(1/n)\sum_{i\in[n]} \mathsf{U}_{ij}^{\alpha}} \triangleq \frac{\boldsymbol{\lambda}_{\boldsymbol{x}}}{\mathsf{Y}}. \tag{37}$$

It is clear that $\bar{\boldsymbol{\lambda}}_{\boldsymbol{x}}$ and $\bar{\mathsf{Y}}$ are an unbiased estimators of $\boldsymbol{\lambda}_{\boldsymbol{x}}$ and Y, respectively. Now, note that we can write $\bar{\boldsymbol{\lambda}}_{\boldsymbol{x}}$ as an average of m i.i.d random variables $\bar{\boldsymbol{x}}_{i_p} = \mathsf{U}_{i_p j}^{\alpha}\boldsymbol{x}_{i_p}$, where i_p is sampled uniformly at random from $[n]$, and included to the set $\mathcal{S}$ in the third step of Algorithm 6 as the p^{th} sample. Similarly, $\bar{\mathsf{Y}}$ can also be written as the average of m i.i.d random variables $\bar{\mathsf{Y}}_{i_p} = \mathsf{U}_{i_p j}^{\alpha}$. Further, notice that $\mathbb{E}\bar{\boldsymbol{x}}_{i_p} = \boldsymbol{\lambda}_{\boldsymbol{x}}$, $\left\|\bar{\boldsymbol{x}}_{i_p}\right\|_2 \leq \mathsf{R}$, and similarly, $\mathbb{E}\bar{\mathsf{Y}}_{i_p} = \mathsf{Y}$, and $\left|\bar{\mathsf{Y}}_{i,p}\right| \leq 1$, for all $p \in \{1, 2, \ldots, m\}$. Next, note that

$$\widehat{\boldsymbol{\mu}}_j = \frac{\bar{\boldsymbol{\lambda}}_{\boldsymbol{x}}}{\bar{\mathsf{Y}}} = \frac{\boldsymbol{\mu}_{\boldsymbol{x}}}{\mathsf{Y}} + \frac{\bar{\boldsymbol{\lambda}}_{\boldsymbol{x}} - \boldsymbol{\lambda}_{\boldsymbol{x}}}{\mathsf{Y}} + \frac{\bar{\boldsymbol{\lambda}}_{\boldsymbol{x}}}{\bar{\mathsf{Y}}}\frac{\mathsf{Y} - \bar{\mathsf{Y}}}{\mathsf{Y}}. \tag{38}$$

Thus, using the triangle inequality we get

$$\left\|\widehat{\boldsymbol{\mu}}_j - \boldsymbol{\mu}_j\right\|_2 \leq \left\|\frac{\bar{\boldsymbol{\lambda}}_{\boldsymbol{x}} - \boldsymbol{\lambda}_{\boldsymbol{x}}}{\mathsf{Y}}\right\|_2 + \left\|\frac{\bar{\boldsymbol{\lambda}}_{\boldsymbol{x}}}{\bar{\mathsf{Y}}}\frac{\mathsf{Y} - \bar{\mathsf{Y}}}{\mathsf{Y}}\right\|_2 \leq \left\|\frac{\bar{\boldsymbol{\lambda}}_{\boldsymbol{x}} - \boldsymbol{\lambda}_{\boldsymbol{x}}}{\mathsf{Y}}\right\|_2 + \mathsf{R}\left\|\frac{\bar{\mathsf{Y}} - \mathsf{Y}}{\mathsf{Y}}\right\|_2, \tag{39}$$

where in the last inequality we have used the fact that $\left\|\widehat{\boldsymbol{\mu}}_j\right\|_2 = \left\|\bar{\boldsymbol{\lambda}}_{\boldsymbol{x}}/\bar{\mathsf{Y}}\right\|_2 \leq \mathsf{R}$. Then, the generalized Hoeffding's inequality in Lemma 4, implies that with probability at least $1 - \delta$,

$$\left\|\bar{\boldsymbol{\lambda}}_{\boldsymbol{x}} - \boldsymbol{\lambda}_{\boldsymbol{x}}\right\|_2^2 \leq \mathsf{R}^2\sqrt{\frac{c}{m}\log\frac{1}{\delta}}, \tag{40}$$

for some $c > 0$, and thus,

$$\left\|\widehat{\boldsymbol{\mu}}_j - \boldsymbol{\mu}_j\right\|_2^2 \leq \frac{4\mathsf{R}^2}{\mathsf{Y}^2}\sqrt{\frac{c}{m}\log\frac{1}{\delta}}. \tag{41}$$

which concludes the proof. $\square$

The following lemma shows that if for a given cluster j we have been able to approximate its center well enough, then Algorithm 2 computes good estimates of the corresponding membership weights with high probability.

Lemma 7 (Estimate of membership given estimated center)**.** *Let $(\mathcal{X}, \mathcal{P})$ be a consistent center-based clustering instance, and recall the definition of $\gamma \in \mathbb{R}_+$ in* (8)*. Assume that for any $j \in [k]$, there exists an estimator $\widehat{\boldsymbol{\mu}}_j$ such that $\left\|\boldsymbol{\mu}_j - \widehat{\boldsymbol{\mu}}_j\right\|_2 \leq \epsilon$ with $\epsilon \leq \gamma$. Then, Algorithm 2 outputs $\widehat{\mathsf{U}}_{ij}$, for $i \in [n]$, such that*

$$0 \leq \mathsf{U}_{ij} - \widehat{\mathsf{U}}_{ij} \leq \eta, \quad \widehat{\mathsf{U}}_{ij} \in \{0, \eta, 2\eta, \ldots, 1\}, \quad \forall i \in [n], \tag{42}$$

for some $\eta \in \mathbb{R}_+$, using $\mathsf{Q} = O\left(\log n/\eta\right)$ queries to the membership-oracle.

Proof of Lemma 7. First, note that since $\mathcal{P}$ is a consistent center-based clustering, we have

$$\mathsf{U}_{\pi_{\boldsymbol{\mu}_j}(i_1)j} \geq \mathsf{U}_{\pi_{\boldsymbol{\mu}_j}(i_2)j} \quad \text{if} \quad i_1 < i_2, i_1, i_2 \in [n]. \tag{43}$$

Indeed, when the elements of $\mathcal{X}$ are sorted in ascending order according to their distance from $\boldsymbol{\mu}_j$, if $\boldsymbol{x}_{i_1}$ is closer to $\boldsymbol{\mu}_j$ than it is to $\boldsymbol{x}_{i_2}$, then $\mathsf{U}_{i_1 j} > \mathsf{U}_{i_2 j}$. Also, since $\left\|\boldsymbol{\mu}_j - \widehat{\boldsymbol{\mu}}_j\right\|_2 \leq \epsilon \leq \gamma$, using (8), this ordering remains the same. Therefore, sorting the elements in $\mathcal{X}$ in ascending order from $\widehat{\boldsymbol{\mu}}_j$ as in the first step of Algorithm 2 gives the same ordering with respect to the true mean. Now, given $\eta \in \mathbb{R}_+$, for each $s \in \{0, 1, 2, \ldots, 1/\eta\}$, in the third step of Algorithm 2 we binary search to find an index ℓ_s such that

$$\ell_s = \underset{i \in [n]}{\mathsf{argmax}}\ \mathsf{U}_{\pi_{\widehat{\boldsymbol{\mu}}_j}(i)j} \geq s\eta. \tag{44}$$

This is done by using $O(\log n/\eta)$ membership-oracle queries. Finally, in the last three steps of Algorithm 2, for each $s \in \{0, 1, 2, \ldots, 1/\eta\}$, and for $i \in \{\ell_s, \ell_s - 1, \ldots, \ell_{(s+1)} + 1\}$, we assign $\widehat{\mathsf{U}}_{\pi_{\widehat{\boldsymbol{\mu}}_j}(i)j} = s\eta$. It is then clear that the estimated memberships satisfy (42), which concludes the proof. $\square$

C Proof of Theorem 1

In this section, we prove Theorem 1. To that end we use the auxiliary results established in the previous section. We start with the following result.

Lemma 8 (Estimate all means)**.** *Let $(\mathcal{X}, \mathcal{P})$ be a consistent center-based clustering instance, recall the definition of $\beta \in (0, 1)$ in* (**??**)*, and let $\delta \in (0, 1)$. Then, with probability at least $1 - \delta$, Algorithm 6 outputs $\widehat{\boldsymbol{\mu}}_j$ such that $\left\|\widehat{\boldsymbol{\mu}}_j - \boldsymbol{\mu}_j\right\|_2 \leq \epsilon$, for all $j \in [k]$, if $m \geq \left(\frac{\mathsf{R}k^\alpha}{\epsilon\beta^\alpha}\right)^4 c \log\frac{k}{\delta}$, for some $c > 0$.*

Proof. Using (**??**) and Hölder's inequality, for any $j \in [k]$ we have,

$$\left(\sum_{i \in [n]} \mathsf{U}_{ij}^\alpha\right)^{1/\alpha} \left(\sum_{i \in [n]} 1^{\alpha/(\alpha-1)}\right)^{(\alpha-1)/\alpha} \geq \sum_{i \in [n]} \mathsf{U}_{ij} \geq \frac{\beta n}{k}, \tag{45}$$

which implies that

$$\left(\sum_{i \in [n]} \mathsf{U}_{ij}^\alpha\right)^{1/\alpha} \geq \frac{\beta n}{k n^{(\alpha-1)/\alpha}} = \frac{\beta n^{1/\alpha}}{k}, \tag{46}$$

and thus

$$\sum_{i \in [n]} \mathsf{U}_{ij}^\alpha \geq \frac{n\beta^\alpha}{k^\alpha}. \tag{47}$$

Therefore,

$$\mathsf{Y} = \min_{j \in [k]} \frac{\sum_{i \in [n]} \mathsf{U}_{ij}^\alpha}{n} \geq \left(\frac{\beta}{k}\right)^\alpha$$

Now, using Lemma 6 and the last result, taking a union bound over all $j \in [k]$, we get

$$\left\|\widehat{\boldsymbol{\mu}}_j - \boldsymbol{\mu}_j\right\|_2 \leq \frac{2\mathsf{R}k^\alpha}{\beta^\alpha}\left(\frac{c}{m}\log\frac{1}{\delta'}\right)^{1/4} \leq \epsilon. \tag{48}$$

with probability $1 - k\delta'$. Rearranging terms and substituting $\delta = k\delta'$, we get the proof of the lemma. $\square$

Proof of Theorem 1. We are now in a position to prove Theorem 1. Using Lemma 8, we can conclude that by taking $m \geq \left(\frac{\mathsf{R}k^\alpha}{\epsilon\beta^\alpha}\right)^4 c\log\frac{k}{\delta}$ in Algorithm 1, which would require km membership-oracle queries, we get $\left\|\widehat{\boldsymbol{\mu}}_j - \boldsymbol{\mu}_j\right\|_2 \leq \epsilon$, for all $j \in [k]$. The time-complexity required to estimate all these means is of order $O(kdm)$. Furthermore, using Lemma 7, using $O(\log n/\eta)$ membership-oracle queries, Algorithm 1 outputs membership estimates such that (42) holds. This requires a time-complexity of order $O(\log n/\eta)$. We note, however, that the membership $\{\widehat{\mathsf{U}}_{ij}\}_{j=1}^k$, for any $i \in [n]$, may not sum up to unity, which is an invalid solution. To fix that in step 7 of Algorithm 1 we add to each $\widehat{\mathsf{U}}_{ij}$ a factor of $\frac{1-\sum_{j=1}^k \widehat{\mathsf{U}}_{ij}}{k}$, and then it is clear that the new estimated membership weights sum up to unity. Furthermore, these updated membership weights satisfy $|\widehat{\mathsf{U}}_{ij} - \mathsf{U}_{ij}| \leq \eta$, for all $i \in [n]$ and $j \in [k]$. Therefore, we have shown that Algorithm 1 is $(\epsilon, \eta, \mathsf{Q})$-solver with probability at least $1 - \delta$, which concludes the proof.

D Proof of Theorem 2

In this section, we prove Theorem 2 using induction.

Base Case: As can be seen from Algorithm 4, in the first step of this algorithm we sample m indices uniformly at random and obtain the multiset $\mathcal{S} \subseteq \mathcal{X}$. Subsequently, we we query U_{ij} for all $i \in \mathcal{S}$ and $j \in [k]$, and then in the third step of the algorithm we choose the cluster t_1 with the highest *membership* value, namely,

$$t_1 = \underset{j\in[k]}{\mathsf{argmax}} \sum_{i\in\mathcal{S}} \mathsf{U}_{ij}^\alpha. \tag{49}$$

Then, in the fourth step of this algorithm we estimate the mean of this cluster by

$$\widehat{\boldsymbol{\mu}}_{t_1} \triangleq \frac{\sum_{i\in\mathcal{S}} \mathsf{U}_{ij}^\alpha \boldsymbol{x}_i}{\sum_{i\in\mathcal{S}} \mathsf{U}_{ij}^\alpha}. \tag{50}$$

We have the following lemma, which is similar to Lemma 6.

Lemma 9 (Guarantees on the largest cluster)**.** *Let $(\mathcal{X}, \mathcal{P})$ be a consistent center-based clustering instance, and let $\delta \in (0, 1)$. With probability at least $1 - \delta/k$, the estimator in* (50) *satisfies*

$$\left\|\widehat{\boldsymbol{\mu}}_{t_1} - \boldsymbol{\mu}_{t_1}\right\|_2 \leq \frac{2\mathsf{R}\left(\frac{c}{m}\log\frac{2k}{\delta}\right)^{1/4}}{\frac{1}{k^\alpha} - \sqrt{\frac{c}{2m}\log\frac{2k}{\delta}}}, \tag{51}$$

where $c > 0$ is an absolute constant.

Proof of Lemma 9. Recall that Lemma 6 tells us that with probability at least $1 - \delta/(2k)$,

$$\left\|\widehat{\boldsymbol{\mu}}_{t_1} - \boldsymbol{\mu}_{t_1}\right\|_2 \leq \frac{2\mathsf{R}}{\mathsf{Y}}\left(\frac{c}{m}\log\frac{2k}{\delta}\right)^{1/4}, \tag{52}$$

where $\mathsf{Y} = (1/n)\sum_{i\in[n]} \mathsf{U}_{it_1}^\alpha$. Now, since t_1 is chosen as the cluster with the maximum membership in the subset $\mathcal{S}$, we will first bound $\widehat{\mathsf{Y}} \triangleq (1/m)\sum_{i\in\mathcal{S}} \mathsf{U}_{it_1}^\alpha$. Notice that $\sum_{i\in\mathcal{S}}\sum_{j=1}^k \mathsf{U}_{ij} = m$, and therefore, using Hölder's inequality we have that for $\alpha > 1$,

$$\left(\sum_{i\in\mathcal{S}}\sum_{j=1}^k \mathsf{U}_{ij}^\alpha\right)^{1/\alpha}\left(\sum_{i\in\mathcal{S}}\sum_{j=1}^k 1^{\alpha/(\alpha-1)}\right)^{(\alpha-1)/\alpha} \geq \sum_{i\in\mathcal{S}}\sum_{j=1}^k \mathsf{U}_{ij} \geq m, \tag{53}$$

which implies that

$$\left(\sum_{i\in\mathcal{S}}\sum_{j=1}^{k}\mathsf{U}_{ij}^{\alpha}\right)^{1/\alpha}\geq\frac{m}{(km)^{(\alpha-1)/\alpha}}=\frac{m^{1/\alpha}}{k^{(\alpha-1)/\alpha}},\tag{54}$$

and therefore,

$$\sum_{i\in\mathcal{S}}\sum_{j=1}^{k}\mathsf{U}_{ij}^{\alpha}\geq\frac{m}{k^{\alpha-1}}.\tag{55}$$

Accordingly, we must have $\sum_{i\in\mathcal{S}}\mathsf{U}_{it_1}\geq\frac{m}{k^{\alpha}}$ which in turn implies that $\widehat{\mathsf{Y}}\geq\frac{1}{k^{\alpha}}$. Next, using Hoeffding's inequality in Lemma 3 we obtain that $|\mathsf{Y}-\widehat{\mathsf{Y}}|\leq\sqrt{\frac{c}{2m}\log\frac{2k}{\delta}}$, with probability at least $1-\delta/(2k)$, and therefore $\mathsf{Y}\geq\widehat{\mathsf{Y}}-\sqrt{\frac{c}{2m}\log\frac{2k}{\delta}}$, which concludes the proof. □

Using the above result and Lemma 7 we obtain the following corollaries.

Corollary 4. *Let $(\mathcal{X},\mathcal{P})$ be a consistent center-based clustering instance, and let $\delta\in(0,1)$. Then, with probability at least $1-\delta/k$, the estimator in (50) satisfies $\left\|\widehat{\boldsymbol{\mu}}_{t_1}-\boldsymbol{\mu}_{t_1}\right\|_2\leq\epsilon$, if $m\geq\frac{c\mathsf{R}^4k^{4\alpha}}{\epsilon^4}\log\frac{2k}{\delta}$. Also, this estimate requires $O\left(\frac{\mathsf{R}^4k^{4\alpha+1}}{\epsilon^4}\log\frac{2k}{\delta}\right)$ membership-oracle queries, and a time-complexity of $O\left(\frac{d\mathsf{R}^4k^{4\alpha}}{\epsilon^4}\log\frac{2k}{\delta}\right)$.*

Proof of Corollary 4. The proof follows from rearranging terms of Lemma 9. The query complexity follows from the fact that we query the membership values U_{ij} for all the $i\in\mathcal{S},j\in[k]$ and the time-complexity follows from the fact that we take the mean of m d-dimensional vectors in order to return the estimator $\widehat{\boldsymbol{\mu}}_{t_1}$. □

Corollary 5. *Let $(\mathcal{X},\mathcal{P})$ be a consistent center-based clustering instance, and recall the definition of $\gamma\in\mathbb{R}_+$ in (8). Assume that there exists an estimator $\widehat{\boldsymbol{\mu}}_{t_1}$ such that $\left\|\boldsymbol{\mu}_{t_1}-\widehat{\boldsymbol{\mu}}_{t_1}\right\|_2\leq\epsilon$ with $\epsilon\leq\gamma$. Then, Algorithm 4 outputs $\widehat{\mathsf{U}}_{it_1}$, for $i\in[n]$, such that*

$$0\leq\mathsf{U}_{it_1}-\widehat{\mathsf{U}}_{it_1}\leq\eta_1,\quad\widehat{\mathsf{U}}_{it_1}\in\{0,\eta_1,2\eta_1,\ldots,1\},\quad\forall i\in[n],\tag{56}$$

for some $\eta\in\mathbb{R}_+$, using $\mathsf{Q}=O\left(\log n/\eta_1\right)$ queries to the membership-oracle.

Proof of Corollary 5. The proof of this lemma follows the same steps as in the proof of Lemma 7. □

Corollaries 4 and 5 show that the base case of our induction is correct.

Induction Hypothesis: We condition on the event that we have been able to estimate $\boldsymbol{\mu}_{t_1},\boldsymbol{\mu}_{t_2},\ldots,\boldsymbol{\mu}_{t_\ell}$ by their corresponding estimators $\widehat{\boldsymbol{\mu}}_{t_1},\widehat{\boldsymbol{\mu}}_{t_2},\ldots,\widehat{\boldsymbol{\mu}}_{t_\ell}$, respectively, such that

$$\left\|\widehat{\boldsymbol{\mu}}_{t_j}-\boldsymbol{\mu}_{t_j}\right\|_2\leq\epsilon,\quad\forall\, j\in[\ell],\tag{57}$$

and further, we have been able to recover U_{it_j}, for all $i\in[n]$ and $j\in[\ell]$, in the sense that

$$0\leq\mathsf{U}_{it_j}-\widehat{\mathsf{U}}_{it_j}\leq\eta_1,\quad\widehat{\mathsf{U}}_{it_j}\in\{0,\eta_1,2\eta_1,\ldots,1\}\quad,\forall\, i\in[n],j\in[\ell].\tag{58}$$

The induction hypothesis states that we have been able to estimate the means of ℓ clusters up to an error of ϵ and subsequently also estimated the memberships of every element in $\mathcal{X}$ to those ℓ clusters such that the estimated memberships are an integral multiple of η_1 and also have a precision error of at most η_1. Given the induction hypothesis, we characterize next the sufficient query complexity and time-complexity required in order to estimate the mean of the $(\ell+1)^{\text{th}}$ cluster and its membership weights.

Inductive Step: Let $\mathsf{Z}_\ell\triangleq\sum_{i\in[n]}\sum_{j\in[\ell]}\mathsf{U}_{it_j}$, and define $\mathcal{X}_s\triangleq\{i\in[n]:\sum_{j\in[\ell]}\widehat{\mathsf{U}}_{it_j}=s\eta_1\}$, for $s\in\{0,1,\ldots,1/\eta_1\}$. In step 10 of Algorithm 4, we sub-sample r indices uniformly at random with replacement from each of the sets $\mathcal{X}_s$, for $s \in \{0, 1, 2, \ldots, 1/\eta_1\}$. Let us denote the multi-set of indices chosen from $\mathcal{X}_s$ by $\mathcal{Y}_s$. Subsequently, in the step 12 of Algorithm 4, for every $s \in \{0, 1, 2, \ldots, 1/\eta_1\}$ and for every element in $\mathcal{Y}_s$, we query the memberships to all the clusters except $t_1, t_2, \ldots, t_\ell$ from the oracle, and set

$$t_{\ell+1} = \underset{j\in[k]\setminus\{t_1,t_2,\ldots,t_\ell\}}{\mathsf{argmax}} \sum_s \frac{|\mathcal{X}_s|}{r} \sum_{i\in\mathcal{Y}_s} \mathsf{U}_{ij}^{\alpha}. \tag{59}$$

Step 13 of Algorithm 4 computes

$$\widehat{\boldsymbol{\mu}}^{t_{\ell+1}} \triangleq \frac{\sum_s \frac{|\mathcal{X}_s|}{r} \sum_{i\in\mathcal{Y}_s} \mathsf{U}_{it_{\ell+1}}^{\alpha} \boldsymbol{x}_i}{\sum_s \frac{|\mathcal{X}_s|}{r} \sum_{i\in\mathcal{Y}_s} \mathsf{U}_{it_{\ell+1}}^{\alpha}}. \tag{60}$$

The following analysis the performance of the estimator in (60). We relegate the proof of this result to the end of this section.

Lemma 10 (Performance of (60))**.** *Let $(\mathcal{X}, \mathcal{P})$ be a consistent center-based clustering instance, and let $\delta \in (0, 1)$. With probability at least $1 - \delta/k$, the estimator in* (60) *satisfies*

$$\left\|\widehat{\boldsymbol{\mu}}_{t_{\ell+1}} - \boldsymbol{\mu}_{t_{\ell+1}}\right\|_2 \leq \frac{2\mathsf{R}\left(\frac{c}{r}\log\frac{4k}{\eta_1\delta}\right)^{1/4}\left(n - \mathsf{Z}_\ell + n\ell\eta_1\right)}{\frac{\left(n-\mathsf{Z}_\ell-n\ell\eta_1\sqrt{\frac{c}{2r}\log\frac{4k}{\eta_1\delta}}\right)^{\alpha}}{n^{\alpha-1}(k-\ell)^{\alpha}} - (n - \mathsf{Z}_\ell - n\ell\eta_1)\sqrt{\frac{c}{2r}\log\frac{4k}{\eta_1\delta}}}, \tag{61}$$

where $c > 0$ is an absolute constant. The query and time-complexity required for evaluating this estimator are of order $O(kr/\eta_1)$ and $O(rd/\eta_1)$, respectively.

Using the above result and Lemma 7 we obtain the following corollaries.

Corollary 6. *Let $(\mathcal{X}, \mathcal{P})$ be a consistent center-based clustering instance, recall the definition of $\beta \in (0, 1)$ in* (**??**)*, and let $\delta \in (0, 1)$. Then, with probability at least $1 - \delta/k$, the estimator in* (60) *satisfies $\left\|\widehat{\boldsymbol{\mu}}_{t_{\ell+1}} - \boldsymbol{\mu}_{t_{\ell+1}}\right\|_2 \leq \epsilon$, if $r \geq \frac{c'\mathsf{R}^4k^{4\alpha}}{\epsilon^4\beta^{4\alpha-4}}\log\frac{4k}{\eta_1\delta}$ and $\eta_1 \leq \frac{1}{k}\left(1 - \frac{\beta}{k}\right)$. Also, this estimate requires $O\left(\frac{\mathsf{R}^4k^{4\alpha+1}}{\epsilon^4\beta^{4\alpha-4}}\log\frac{4k}{\eta_1\delta}\right)$ membership-oracle queries, and a time-complexity of $O\left(\frac{\mathsf{R}^4k^{4\alpha}}{\epsilon^4\beta^{4\alpha-4}}\log\frac{4k}{\eta_1\delta}\right)$.*

Proof of Corollary 6. Using (**??**) and the fact that $n - \mathsf{Z}_\ell = \sum_{i\in[n]}\sum_{j\in[k]\setminus[\ell]} \mathsf{U}_{it_j}$, we have the following upper and lower bound

$$\frac{n - \mathsf{Z}_\ell}{k - \ell} \geq \frac{\beta n}{k} \quad \text{and} \quad n - \mathsf{Z}_\ell \leq n - \frac{\beta n}{k}, \tag{62}$$

which follows from the fact that the average *membership size* of the any $k - \ell$ clusters must be larger than the membership size of the smallest cluster. Thus, if $\eta_1 \leq \frac{1}{k}\left(1 - \frac{\beta}{k}\right)$, as claimed in the statement of the lemma, we have $n\ell\eta_1 \leq n - \mathsf{Z}_\ell$. With the chosen values of η_1 and r, and the fact that $\ell < k$, we get $n\ell\eta_1\sqrt{\frac{c}{2r}\log\frac{4k}{\eta_1\delta}} = o(n - \mathsf{Z}_\ell)$. Therefore,

$$\left\|\widehat{\boldsymbol{\mu}}_{t_{\ell+1}} - \boldsymbol{\mu}_{t_{\ell+1}}\right\|_2 \leq \frac{4\mathsf{R}\left(\frac{c}{r}\log\frac{4k}{\eta_1\delta}\right)^{1/4}(n - \mathsf{Z}_\ell)}{\frac{(n-\mathsf{Z}_\ell)^{\alpha}}{n^{\alpha-1}(k-\ell)^{\alpha}} - 2\sqrt{\frac{c}{2r}\log\frac{4k}{\eta_1\delta}}(n - \mathsf{Z}_\ell)} \tag{63}$$

$$= \frac{4\mathsf{R}\left(\frac{c}{r}\log\frac{4k}{\eta_1\delta}\right)^{1/4}}{\frac{(n-\mathsf{Z}_\ell)^{\alpha-1}}{n^{\alpha-1}(k-\ell)^{\alpha}} - 2\sqrt{\frac{c}{2r}\log\frac{4k}{\eta_1\delta}}} \tag{64}$$

$$\leq \frac{4\mathsf{R}\left(\frac{c}{r}\log\frac{4k}{\eta_1\delta}\right)^{1/4}}{\frac{\beta^{\alpha-1}}{k^{\alpha}(k-\ell)} - 2\sqrt{\frac{c}{2r}\log\frac{4k}{\eta_1\delta}}}. \tag{65}$$

Again, for the chosen value of r, it is clear that $\sqrt{\frac{c}{2r}\log\frac{4k}{\eta_1\delta}} = o\left(\frac{\beta^{\alpha-1}}{k^{\alpha}(k-\ell)}\right)$, and thus we get that $\left\|\widehat{\boldsymbol{\mu}}_{t_{\ell+1}} - \boldsymbol{\mu}_{t_{\ell+1}}\right\|_2 \leq \epsilon$, with probability at least $1 - \delta/k$. $\square$

Corollary 7. *Let $(\mathcal{X}, \mathcal{P})$ be a consistent center-based clustering instance, and recall the definition of $\gamma \in \mathbb{R}_+$ in* (8)*. Assume that there exists an estimator $\widehat{\boldsymbol{\mu}}_{t_{\ell+1}}$ such that $\left\|\boldsymbol{\mu}_{t_{\ell+1}} - \widehat{\boldsymbol{\mu}}_{t_{\ell+1}}\right\|_2 \leq \epsilon$ with $\epsilon \leq \gamma$. Then, Algorithm 4 outputs $\widehat{\mathsf{U}}_{it_{\ell+1}}$, for $i \in [n]$, such that*

$$0 \leq \mathsf{U}_{it_{\ell+1}} - \widehat{\mathsf{U}}_{it_{\ell+1}} \leq \eta_1, \quad \widehat{\mathsf{U}}_{it_{\ell+1}} \in \{0, \eta_1, 2\eta_1, \ldots, 1\}, \quad \forall i \in [n], \tag{66}$$

for some $\eta \in \mathbb{R}_+$, using $\mathsf{Q} = O\left(\log n/\eta_1\right)$ queries to the membership-oracle.

Proof of Corollary 7. The proof of this lemma follows the same steps as in the proof of Lemma 7. □

<u>Proof of Theorem 2.</u> We are now in a position to proof Theorem 2. To that end, we use our induction mechanism. Specifically, Corollaries 4 and 5 prove the base case for the first cluster t_1. Subsequently, Corollaries 6 and 7 prove the induction step after taking a union bound over all clusters and using $\eta_1 = \frac{1}{k}\left(1 - \frac{\beta}{k}\right)$. Finally, we can use Lemma 7 in order to estimate the memberships $\mathsf{U}_{ij}\ \forall\ i \in [n], j \in [k]$ up to a precision of η_2 using an addition query and time-complexity of $O(k \log n/\eta_2)$.

It is left to prove Lemma 10.

Proof of Lemma 10. Let $\Sigma \triangleq \{0, 1, \ldots, 1/\eta_1\}$. We have

$$\boldsymbol{\mu}_{t_{\ell+1}} = \frac{\sum_{i\in[n]} \mathsf{U}^{\alpha}_{it_{\ell+1}} \boldsymbol{x}_i}{\sum_{i\in[n]} \mathsf{U}^{\alpha}_{it_{\ell+1}}} = \frac{\sum_{s\in\Sigma}\sum_{i\in\mathcal{X}_s} \mathsf{U}^{\alpha}_{it_{\ell+1}} \boldsymbol{x}_i}{\sum_{s\in\Sigma}\sum_{i\in\mathcal{X}_s} \mathsf{U}^{\alpha}_{it_{\ell+1}}} \triangleq \frac{\sum_{s\in\Sigma} \boldsymbol{\lambda}_s}{\sum_{s\in\Sigma} \mathsf{Y}_s}, \tag{67}$$

and that

$$\widehat{\boldsymbol{\mu}}_{t_{\ell+1}} = \frac{\sum_{s\in\Sigma} \frac{|\mathcal{X}_s|}{r} \sum_{i\in\mathcal{Y}_s} \mathsf{U}^{\alpha}_{it_{\ell+1}} \boldsymbol{x}_i}{\sum_{s\in\Sigma} \frac{|\mathcal{X}_s|}{r} \sum_{i\in\mathcal{Y}_s} \mathsf{U}^{\alpha}_{it_{\ell+1}}} \triangleq \frac{\sum_{s\in\Sigma} \bar{\boldsymbol{\lambda}}_s}{\sum_{s\in\Sigma} \bar{\mathsf{Y}}_s}. \tag{68}$$

Now, note that we can write $\bar{\boldsymbol{\lambda}}_s$ as an average of r i.i.d random variables $\bar{\boldsymbol{x}}_{s,i_p} \triangleq |\mathcal{X}_s|\, \mathsf{U}^{\alpha}_{i_p t_{\ell+1}} \boldsymbol{x}_{i_p}$, where i_p is sampled uniformly at random from $[n]$, and included to the set $\mathcal{Y}_s$ in the step 9 of Algorithm 4 as the p^{th} sample. Similarly, $\bar{\mathsf{Y}}_s$ can also be written as the average of r i.i.d random variables $\bar{\mathsf{Y}}_{s,i_p} = |\mathcal{X}_s|\, \mathsf{U}^{\alpha}_{i_p t_{\ell+1}}$. Therefore, it is evident that $\mathbb{E}\bar{\boldsymbol{x}}_{s,i_p} = \boldsymbol{\lambda}_s$ and $\mathbb{E}\bar{\mathsf{Y}}_{s,i_p} = \mathsf{Y}_s$ for all $p \in [r]$. This implies that the numerator and denominator of (68) are both unbiased. Now, note that

$$\widehat{\boldsymbol{\mu}}_{t_{\ell+1}} = \frac{\sum_{s\in\Sigma} \bar{\boldsymbol{\lambda}}_s}{\sum_{s\in\Sigma} \bar{\mathsf{Y}}_s} = \frac{\sum_{s\in\Sigma} \boldsymbol{\lambda}_s}{\sum_{s\in\Sigma} \mathsf{Y}_s} + \frac{\sum_{s\in\Sigma} \bar{\boldsymbol{\lambda}}_s - \boldsymbol{\lambda}_s}{\sum_{s\in\Sigma} \mathsf{Y}_s} + \frac{\sum_{s\in\Sigma} \bar{\boldsymbol{\lambda}}_s}{\sum_{s\in\Sigma} \bar{\mathsf{Y}}_s} \frac{\sum_{s\in\Sigma} \mathsf{Y}_s - \bar{\mathsf{Y}}_s}{\sum_{s\in\Sigma} \mathsf{Y}_s}. \tag{69}$$

Thus, using the triangle inequality we get

$$\left\|\widehat{\boldsymbol{\mu}}_{t_{\ell+1}} - \boldsymbol{\mu}_{t_{\ell+1}}\right\|_2 \leq \sum_{s\in\Sigma} \left\|\frac{\bar{\boldsymbol{\lambda}}_s - \boldsymbol{\lambda}_s}{\sum_{s\in\Sigma} \mathsf{Y}_s}\right\|_2 + \mathsf{R} \sum_{s\in\Sigma} \left\|\frac{\mathsf{Y}_s - \bar{\mathsf{Y}}_s}{\sum_{s\in\Sigma} \mathsf{Y}_s}\right\|_2, \tag{70}$$

where the last inequality follows from the fact that $\left\|\widehat{\boldsymbol{\mu}}_{t_{\ell+1}}\right\|_2 = \left\|\frac{\sum_{s\in\Sigma} \bar{\boldsymbol{\lambda}}_s}{\sum_{s\in\Sigma} \bar{\mathsf{Y}}_s}\right\|_2 \leq \mathsf{R}$.

Next, we note that for $i \in \mathcal{Y}_s$, we have

$$\mathsf{U}^{\alpha}_{it_{\ell+1}} \leq (1 - \sum_{j\in[\ell]} \mathsf{U}_{it_j})^{\alpha} \leq (1 - \sum_{j\in[\ell]} \widehat{\mathsf{U}}_{it_j})^{\alpha} = (1 - s\eta_1)^{\alpha}, \tag{71}$$

where we have used the induction hypothesis in (58). Also, using Lemmas 3 and 4, for all $s \in \Sigma$, with probability at least $1 - \delta/2k$,

$$\begin{aligned} \left|\left|\bar{\boldsymbol{\lambda}}_s - \boldsymbol{\lambda}_s\right|\right|_2 &\leq \mathsf{R}|\mathcal{X}_s|(1 - s\eta_1)^{\alpha} \left(\frac{c}{r} \log \frac{4k}{\eta_1 \delta}\right)^{1/4} \\ &\leq \mathsf{R}|\mathcal{X}_s|(1 - s\eta_1) \left(\frac{c}{r} \log \frac{4k}{\eta_1 \delta}\right)^{1/4}, \end{aligned} \tag{72}$$

and

$$\begin{aligned}\left|\bar{\mathsf{Y}}_s - \mathsf{Y}_s\right| &\leq |\mathcal{X}_s|(1-s\eta_1)^{\alpha}\sqrt{\frac{c}{r}\log\frac{4k}{\eta_1\delta}} \\ &\leq |\mathcal{X}_s|(1-s\eta_1)\sqrt{\frac{c}{2r}\log\frac{4k}{\eta_1\delta}}. \end{aligned} \tag{73}$$

Using the induction hypothesis in (58) once again, we have

$$\mathsf{Z}_\ell = \sum_{s\in\Sigma}\sum_{i\in\mathcal{X}_s}\sum_{j\in[\ell]}\mathsf{U}_{it_j} \leq \sum_{s\in\Sigma}\sum_{i\in\mathcal{X}_s}\left(\ell\eta_1 + \sum_{j\in[\ell]}\widehat{\mathsf{U}}_{it_j}\right) \leq \sum_{s\in\Sigma}|\mathcal{X}_s|\, s\eta_1 + n\ell\eta_1, \tag{74}$$

and thus

$$\sum_{s\in\Sigma}|\mathcal{X}_s|(1-s\eta_1) \leq n - \mathsf{Z}_\ell + n\ell\eta_1. \tag{75}$$

Next, we lower bound $\sum_{s\in\Sigma}\mathsf{Y}_s$. To that end, in light of (70), it is suffice to bound $\sum_{s\in\Sigma}\bar{\mathsf{Y}}_s$. Using Hölder's inequality we have

$$\begin{aligned}&\Bigg(\sum_{i\in\mathcal{Y}_s}\sum_{j\in[k]\setminus\{t_1,t_2,\ldots,t_\ell\}}\mathsf{U}_{ij}^{\alpha}\Bigg)^{1/\alpha}\Bigg(\sum_{i\in\mathcal{Y}_s}\sum_{j\in[k]\setminus\{t_1,t_2,\ldots,t_\ell\}}1^{\alpha/(\alpha-1)}\Bigg)^{(\alpha-1)/\alpha} \\ &\qquad\geq \sum_{i\in\mathcal{Y}_s}\sum_{j\in[k]\setminus\{t_1,t_2,\ldots,t_\ell\}}\mathsf{U}_{ij}, \end{aligned} \tag{76}$$

which implies that

$$\Bigg(\sum_{i\in\mathcal{Y}_s}\sum_{j\in[k]\setminus\{t_1,t_2,\ldots,t_\ell\}}\mathsf{U}_{ij}^{\alpha}\Bigg)^{1/\alpha} \geq \frac{\sum_{i\in\mathcal{Y}_s}\sum_{j\in[k]\setminus\{t_1,t_2,\ldots,t_\ell\}}\mathsf{U}_{ij}}{(r(k-\ell))^{(\alpha-1)/\alpha}}, \tag{77}$$

and therefore,

$$\sum_{i\in\mathcal{Y}_s}\sum_{j\in[k]\setminus\{t_1,t_2,\ldots,t_\ell\}}\mathsf{U}_{ij}^{\alpha} \geq \frac{\left(\sum_{i\in\mathcal{Y}_s}\sum_{j\in[k]\setminus\{t_1,t_2,\ldots,t_\ell\}}\mathsf{U}_{ij}\right)^{\alpha}}{(r(k-\ell))^{(\alpha-1)}}. \tag{78}$$

To further lower bound the r.h.s. of (78), we use the power means inequality and ger

$$\frac{\sum_{s\in\Sigma}\frac{|\mathcal{X}_s|}{r}\left(\sum_{i\in\mathcal{Y}_s}\sum_{j\in[k]\setminus\{t_1,t_2,\ldots,t_\ell\}}\mathsf{U}_{ij}\right)^{\alpha}}{\sum_{s'\in\Sigma}\frac{|\mathcal{X}_{s'}|}{r}} \geq \left(\frac{\sum_{s\in\Sigma}\frac{|\mathcal{X}_s|}{r}\sum_{i\in\mathcal{Y}_s}\sum_{j\in[k]\setminus\{t_1,t_2,\ldots,t_\ell\}}\mathsf{U}_{ij}}{\sum_{s'\in\Sigma}\frac{|\mathcal{X}_{s'}|}{r}}\right)^{\alpha}. \tag{79}$$

Thus, the fact that $\sum_{s\in\Sigma}\frac{|\mathcal{X}_s|}{r} = \frac{n}{r}$, combined with (78) and (79), imply that

$$\sum_{s\in\Sigma}\frac{|\mathcal{X}_s|}{r}\sum_{i\in\mathcal{Y}_s}\sum_{j\in[k]\setminus\{t_1,t_2,\ldots,t_\ell\}}\mathsf{U}_{ij}^{\alpha} \geq \frac{\left(\sum_{s\in\Sigma}\frac{|\mathcal{X}_s|}{r}\sum_{i\in\mathcal{Y}_s}\sum_{j\in[k]\setminus\{t_1,t_2,\ldots,t_\ell\}}\mathsf{U}_{ij}\right)^{\alpha}}{(n(k-\ell))^{(\alpha-1)}} \tag{80}$$

$$= \frac{\left(n - \sum_{s\in\Sigma}\frac{|\mathcal{X}_s|}{r}\sum_{i\in\mathcal{Y}_s}\sum_{j\in\{t_1,t_2,\ldots,t_\ell\}}\mathsf{U}_{ij}\right)^{\alpha}}{(n(k-\ell))^{(\alpha-1)}}. \tag{81}$$

We next upper bound the term inside the brackets at the r.h.s. of (81). To that end, for a given s, we define the random variables

$$\mathsf{H}_{s,i_p} \triangleq |\mathcal{X}_s|\left(\sum_{j\in\{t_1,t_2,\ldots,t_\ell\}}\mathsf{U}_{i_pj} - s\eta_1\right), \tag{82}$$

where where i_p is sampled uniformly at random from $[n]$, and included to the set $\mathcal{Y}_s$ in the step 9 of Algorithm 4 as the p^{th} sample. With this definition, it is evident that $\frac{|\mathcal{X}_s|}{r}\sum_{i\in\mathcal{Y}_s}(\sum_{j\in\{t_1,t_2,\ldots,t_\ell\}}\mathsf{U}_{ij} - s\eta_1)$ can be written as the average of these r i.i.d random variables, namely,

$$\frac{|\mathcal{X}_s|}{r}\sum_{i\in\mathcal{Y}_s}\left(\sum_{j\in\{t_1,t_2,\ldots,t_\ell\}}\mathsf{U}_{ij} - s\eta_1\right) = \frac{1}{r}\sum_{i\in\mathcal{Y}_s}\mathsf{H}_{s,i}. \tag{83}$$

Note that $\mathbb{E}\mathsf{H}_{s,i_p} = \sum_{i\in\mathcal{X}_s}\sum_{j\in\{t_1,t_2,\ldots,t_\ell\}}\mathsf{U}_{ij} - |\mathcal{X}_s|s\eta_1$ and $\left|\mathsf{H}_{s,i_p}\right| \leq \ell\eta_1$ for all $i_p \in [r]$. For simplicity of notation we define $\mathsf{Z}_{s\ell} \triangleq \sum_{i\in\mathcal{X}_s}\sum_{j\in\{t_1,t_2,\ldots,t_\ell\}}\mathsf{U}_{ij}$. Then, using Hoeffding's inequality in Lemma 3, we have with probability at least $1-\delta/4k$,

$$\begin{aligned}\left|\frac{|\mathcal{X}_s|}{r}\sum_{i\in\mathcal{Y}_s}\left(\sum_{j\in\{t_1,t_2,\ldots,t_\ell\}}\mathsf{U}_{ij} - s\eta_1\right) - \mathsf{Z}_{s\ell} + |\mathcal{X}_s|s\eta_1\right| \\ \leq |\mathcal{X}_s|\,\ell\eta_1\sqrt{\frac{c}{2r}\log\frac{4k}{\eta_1\delta}},\end{aligned} \tag{84}$$

and thus,

$$\frac{|\mathcal{X}_s|}{r}\sum_{i\in\mathcal{Y}_s}\sum_{j\in\{t_1,t_2,\ldots,t_\ell\}}\mathsf{U}_{ij} \leq \max\left(|\mathcal{X}_s|\,,\mathsf{Z}_{s\ell} + |\mathcal{X}_s|\,\ell\eta_1\sqrt{\frac{c}{2r}\log\frac{4k}{\eta_1\delta}}\right) \tag{85}$$

$$= \mathsf{Z}_{s\ell} + \max\left(|\mathcal{X}_s| - \mathsf{Z}_{s\ell}, |\mathcal{X}_s|\,\ell\eta_1\sqrt{\frac{c}{2r}\log\frac{4k}{\eta_1\delta}}\right). \tag{86}$$

Summing (86) over $s\in\Sigma$ and using the fact that $\sum_{s\in\Sigma}\mathsf{Z}_{s\ell} = \mathsf{Z}_\ell$, we obtain

$$n - \sum_{s\in\Sigma}\frac{|\mathcal{X}_s|}{r}\sum_{i\in\mathcal{Y}_s}\sum_{j\in\{t_1,t_2,\ldots,t_\ell\}}\mathsf{U}_{ij} \geq n - \mathsf{Z}_\ell - \min\left(n-\mathsf{Z}_\ell, n\ell\eta_1\sqrt{\frac{c}{2r}\log\frac{4k}{\eta_1\delta}}\right). \tag{87}$$

Substituting the last inequality in (81), we finally get

$$\sum_{s\in\Sigma}\frac{|\mathcal{X}_s|}{r}\sum_{i\in\mathcal{Y}_s}\sum_{j\in[k]\setminus\{t_1,t_2,\ldots,t_\ell\}}\mathsf{U}_{ij}^\alpha \geq \frac{\left(n-\mathsf{Z}_\ell - n\ell\eta_1\sqrt{\frac{c}{2r}\log\frac{4k}{\eta_1\delta}}\right)^\alpha}{(n(k-\ell))^{(\alpha-1)}}. \tag{88}$$

Next, recall that the index of the $(\ell+1)^{\text{th}}$ cluster is chosen as

$$t_{\ell+1} = \underset{j\in[k]\setminus\{t_1,t_2,\ldots,t_\ell\}}{\mathsf{argmax}}\sum_{s\in\Sigma}\frac{|\mathcal{X}_s|}{r}\sum_{i\in\mathcal{Y}_s}\mathsf{U}_{ij}^\alpha, \tag{89}$$

and therefore,

$$\sum_{s\in\Sigma}\bar{\mathsf{Y}}_s = \sum_{s\in\Sigma}\frac{|\mathcal{X}_s|}{r}\sum_{i\in\mathcal{Y}_s}\mathsf{U}_{it_{\ell+1}}^\alpha \tag{90}$$

$$\geq \frac{\left(n-\mathsf{Z}_\ell - n\ell\eta_1\sqrt{\frac{c}{2r}\log\frac{4k}{\eta_1\delta}}\right)^\alpha}{n^{\alpha-1}(k-\ell)^\alpha}. \tag{91}$$

Combining (73), (75), and (91), we get a lower bound on $\sum_{s\in\Sigma}\mathsf{Y}_s$ as follows

$$\sum_{s\in\Sigma}\mathsf{Y}_s \geq \frac{\left(n-\mathsf{Z}_\ell - n\ell\eta_1\sqrt{\frac{c}{2r}\log\frac{4k}{\eta_1\delta}}\right)^\alpha}{n^{\alpha-1}(k-\ell)^\alpha} - (n-\mathsf{Z}_\ell - n\ell\eta_1)\sqrt{\frac{c}{2r}\log\frac{4k}{\eta_1\delta}}. \tag{92}$$

Algorithm 7 MEMBERSHIP2$(\mathcal{X}, \widehat{\boldsymbol{\mu}}_{t_1}, \alpha)$ Estimate the memberships U_{it_2} for all $i \in [n]$ given an estimated mean $\widehat{\boldsymbol{\mu}}_{t_1}$

Input: $\mathcal{O}_{\mathsf{fuzzy}}$.

Sort the elements $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ in ascending order according to $\left\|\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}}_{t_1}\right\|_2$. Denote the resultant permutation of $[n]$ corresponding to the sorted elements by $\pi_{\widehat{\boldsymbol{\mu}}_{t_1}}$.
Set $\eta_1 \triangleq 1 - \mathsf{U}_{\pi_{\widehat{\boldsymbol{\mu}}_{t_1}}(n)t_1}$ and $p_{\eta_1} \triangleq n$.
Query $\mathcal{O}_{\mathsf{fuzzy}}(\pi_{\widehat{\boldsymbol{\mu}}_{t_1}}(n), t_1)$ to obtain $\mathsf{U}_{\pi_{\widehat{\boldsymbol{\mu}}_{t_1}}(n)t_1}$. Set $\mathsf{U}_{\pi_{\widehat{\boldsymbol{\mu}}_{t_1}}(n),t_2} = 1 - \mathsf{U}_{\pi_{\widehat{\boldsymbol{\mu}}_{t_1}}(n)t_1}$.
Initialize $\mathcal{P}_1, \mathcal{P}_2 = \phi$
for $q = 2, 3, \ldots, 3\log n$ **do**
  Initialize $\mathcal{X}_q = \phi$ and set $\eta_q = \frac{\eta_{q-1}}{2}$
  Find $p'_{\eta_q} = \mathsf{argmin}_{i\in[n]} \mathsf{U}_{\pi_{\widehat{\boldsymbol{\mu}}_{t_1}}(i),t_2} \geq \eta_q$ using BINARYSEARCH2$(\mathcal{X}, \pi_{\widehat{\boldsymbol{\mu}}_{t_1}}, \eta_q)$.
  if $|p'_{\eta_q} - p_{\eta_{q-1}}| \geq \log n$ **then**
    Set $p_{\eta_q} = p'_{\eta_q}$
    for $i = p_{\eta_q}, p_{\eta_q}+1, \ldots, p_{\eta_{q-1}} - 1$ **do**
      Set $\mathcal{X}_q = \mathcal{X}_q \cup \pi_{\widehat{\boldsymbol{\mu}}_{t_1}}(i)$ and set $\widehat{\mathsf{U}}_{\pi_{\widehat{\boldsymbol{\mu}}_{t_1}}(i),t_2} = 1 - \mathsf{U}_{\pi_{\widehat{\boldsymbol{\mu}}_{t_1}}(p_{\eta_q})t_1}$
    end for
  else
    Query $\mathcal{O}_{\mathsf{fuzzy}}(\pi_{\widehat{\boldsymbol{\mu}}_{t_1}}(\min(0, p_{\eta_{q-1}} - \log n - 1)), t_1)$ and obtain the membership $\mathsf{U}_{\pi_{\widehat{\boldsymbol{\mu}}_{t_1}}(\min(0,p_{\eta_{q-1}}-\log n-1)),t_1}$.
    Set $\eta_q = \mathsf{U}_{\pi_{\widehat{\boldsymbol{\mu}}_{t_1}}(\min(0,p_{\eta_{q-1}}-\log n-1)),t_2}$ and $p_{\eta_q} = \min(0, p_{\eta_{q-1}} - 1 - \log n)$.
    for $i = p_{\eta_q}, p_{\eta_q}+1, \ldots, p_{\eta_{q-1}} - 1$ **do**
      Query $\mathcal{O}_{\mathsf{fuzzy}}(\pi_{\widehat{\boldsymbol{\mu}}_{t_1}}(i), t_1)$ to obtain $\mathsf{U}_{\pi_{\widehat{\boldsymbol{\mu}}_{t_1}}(i)t_1}$.
      Set $\widehat{\mathsf{U}}_{\pi_{\widehat{\boldsymbol{\mu}}_{t_1}}(i),t_2} = 1 - \mathsf{U}_{\pi_{\widehat{\boldsymbol{\mu}}_{t_1}}(i)t_1}$ and set $\mathcal{X}'_q = \mathcal{X}'_q \cup \pi_{\widehat{\boldsymbol{\mu}}_{t_1}}(i)$
    end for
  end if
end for
for $i = 0, 1, \ldots, \eta_{3\log n} - 1$ **do**
  Set $\widehat{\mathsf{U}}_{\pi_{\widehat{\boldsymbol{\mu}}_{t_1}}(i),t_2} = 0$ and set $\mathcal{P}_1 = \mathcal{P}_1 \cup \pi_{\widehat{\boldsymbol{\mu}}_{t_1}}(i)$
end for
Set $\mathcal{P}_2 = \cup_q \mathcal{X}'_q$.
Return $\mathcal{P}_1, \mathcal{P}_2, \mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_q$ and $\widehat{\mathsf{U}}_{it_2}$ for all $i \in [n]$.

Finally, combining (70), (72), and (92), we get

$$\left\|\widehat{\boldsymbol{\mu}}_{t_{\ell+1}} - \boldsymbol{\mu}_{t_{\ell+1}}\right\|_2 \leq \sum_{s\in\Sigma}\left\|\frac{\bar{\boldsymbol{\lambda}}_s - \boldsymbol{\lambda}_s}{\sum_{s\in\Sigma}\mathsf{Y}_s}\right\|_2 + \mathsf{R}\sum_{s\in\Sigma}\left\|\frac{\mathsf{Y}_s - \bar{\mathsf{Y}}_s}{\sum_{s\in\Sigma}\mathsf{Y}_s}\right\|_2 \tag{93}$$

$$\leq \frac{2\mathsf{R}\left(\frac{c}{r}\log\frac{4k}{\eta_1\delta}\right)^{1/4}\left(n - \mathsf{Z}_\ell + n\ell\eta_1\right)}{\frac{\left(n-\mathsf{Z}_\ell-n\ell\eta_1\sqrt{\frac{c}{2r}\log\frac{4k}{\eta_1\delta}}\right)^{\alpha}}{n^{\alpha-1}(k-\ell)^{\alpha}} - (n - \mathsf{Z}_\ell - n\ell\eta_1)\sqrt{\frac{c}{2r}\log\frac{4k}{\eta_1\delta}}}, \tag{94}$$

with probability at least $1 - \delta/k$, which concludes the proof. □

E Proof of Theorem 3

To prove Theorem 3 we will establish some preliminary results.

E.1 Auxiliary Lemmata

We start with the following result which shows that given a good estimate for the larger cluster among the two, we can approximate the membership weights of the smaller cluster reliably. This is done in Procedure 7.

Lemma 11. *Let $(\mathcal{X}, \mathcal{P})$ be a consistent center-based clustering instance, recall the definition of $\gamma \in \mathbb{R}_+$ in* (8)*, and let $k = 2$. Assume that there exists an estimator $\widehat{\boldsymbol{\mu}}_{t_1}$ such that $\left\|\boldsymbol{\mu}_{t_1} - \widehat{\boldsymbol{\mu}}_{t_1}\right\|_2 \leq \epsilon$ with $\epsilon \leq \gamma$. Then, Procedure 7 outputs $\widehat{\mathsf{U}}_{it_2}$, for $i \in [n]$, such that*

$$\widehat{\mathsf{U}}_{it_2} \in \mathcal{A} \subset \{\mathsf{U}_{it_2} : i \in [n]\}, \tag{95}$$

Algorithm 8 BINARYSEARCH2($\mathcal{X}, \pi, x$): Search for the minimum index i such that $1 - \mathsf{U}_{\pi(i)t_1} \geq x$

Input: $\mathcal{O}_{\mathsf{fuzzy}}$.

Set $\mathsf{low} = 1$ and $\mathsf{high} = n$.
while $\mathsf{low} \neq \mathsf{high}$ **do**
  Set $\mathsf{mid} = \lfloor (\mathsf{low} + \mathsf{high})/2 \rfloor$.
  Query $\mathcal{O}_{\mathsf{fuzzy}}(\pi(\mathsf{mid})t_1)$ to obtain $\mathsf{U}_{\pi(\mathsf{mid})t_1}$.
  if $\mathsf{U}_{\pi(\mathsf{mid})t_1} \geq 1 - x$ **then**
    Set $\mathsf{low} = \mathsf{mid} + 1$
  else
    Set $\mathsf{high} = \mathsf{mid}$
  end if
end while
Return low.

with $|\mathcal{A}| = O(\log^2 n)$, and

$$\max_{i \in [n]: \widehat{\mathsf{U}}_{it_2} = x} \mathsf{U}_{it_2} \leq \max\left(\frac{\max_{i \in [n]} \mathsf{U}_{it_2}}{n^3}, \min_{i \in [n]: \widehat{\mathsf{U}}_{it_2} = x} 2\mathsf{U}_{it_2} \right), \tag{96}$$

for all $x \in \mathcal{A}$, and for some $\eta \in \mathbb{R}_+$, using $\mathsf{Q} = O(\log^2 n)$ queries to the membership-oracle, and time-complexity $O(\log^2 n)$.

Proof of Lemma 11. First, note that since $\mathcal{P}$ is a consistent center-based clustering, we have

$$\mathsf{U}_{\sigma_{\boldsymbol{\mu}_{t_1}}(r_1)t_2} \leq \mathsf{U}_{\sigma_{\boldsymbol{\mu}_{t_1}}(r_2)t_2} \quad \text{if} \quad r_1 < r_2,\ r_1, r_2 \in [n]. \tag{97}$$

Indeed, when the elements of $\mathcal{X}$ are sorted in ascending order according to their distance from $\boldsymbol{\mu}_{t_1}$, if $\boldsymbol{x}_i$ is closer to $\boldsymbol{\mu}_{t_1}$ than it is to $\boldsymbol{x}_j$, then $\mathsf{U}_{it_1} \geq \mathsf{U}_{jt_1}$ and thus $\mathsf{U}_{it_2} \leq \mathsf{U}_{jt_2}$. Also, since $\left\| \boldsymbol{\mu}_{t_1} - \widehat{\boldsymbol{\mu}}_{t_1} \right\|_2 \leq \epsilon \leq \gamma$, using (8), this ordering remains the same. Therefore, sorting the elements in $\mathcal{X}$ in ascending order from $\widehat{\boldsymbol{\mu}}_{t_1}$ as in the first step of Procedure 7 gives the same ordering with respect to the true mean. Now, given $\eta \in \mathbb{R}_+$, we search for the index

$$p'_\eta \triangleq \mathsf{argmin}_{j \in [n]} \mathbb{1}\left[\mathsf{U}_{\pi_{\widehat{\boldsymbol{\mu}}_{t_1}}(j)t_2} \geq \eta \right], \tag{98}$$

which can be done by using the binary search routine in Procedure 8, which ask for at most $O(\log n)$ membership-oracle queries. We will do this step for $\eta_1, \eta_2, \ldots, \eta_{3\log n}$, as described in Algorithm 7. The values of $\{\eta_i\}$ are chosen as follows. We initialize $\eta_1 = \mathsf{U}_{\pi_{\widehat{\boldsymbol{\mu}}_{t_1}}(n)t_2}$ and $p_{\eta_1} \triangleq n$, and update the other values of η_i's recursively as follows. Let $\mathcal{V} \triangleq \{1, 2, \ldots, 3\log n\}$. For each $q \in \mathcal{V} \setminus 1$, we first set $\eta_q = \eta_{q-1}/2$ and subsequently, if $|p'_{\eta_q} - p_{\eta_{q-1}}| \geq \log n$, then η_q remains unchanged and we set $p_{\eta_q} = p'_{\eta_q}$. Otherwise, if $|p'_{\eta_q} - p_{\eta_{q-1}}| < \log n$, then we update both η_q and p_{η_q} as follows:

$$p_{\eta_q} = \min(0, p_{\eta_{q-1}} - 1 - \log n) \tag{99}$$

$$\eta_q = \mathsf{U}_{\pi_{\widehat{\boldsymbol{\mu}}_{t_1}}(p_{\eta_q})t_2} \tag{100}$$

For each value of $q \in \mathcal{V}$, we initialize two sets $\mathcal{X}_q, \mathcal{X}'_q = \phi$. If $|p'_{\eta_q} - p_{\eta_{q-1}}| \geq \log n$, then we update $\mathcal{X}_q = \{\pi_{\widehat{\boldsymbol{\mu}}_{t_1}}(i) : p_{\eta_q} \leq i \leq p_{\eta_{q-1}} - 1, i \in [n]\}$ and if $|p'_{\eta_q} - p_{\eta_{q-1}}| < \log n$, we update $\mathcal{X}'_q = \{\pi_{\widehat{\boldsymbol{\mu}}_{t_1}}(i) : p_{\eta_q} \leq i \leq p_{\eta_{q-1}} - 1, i \in [n]\}$. It is clear that $\eta_q \leq \eta_{q-1}/2$ and therefore, we must have $\eta_{3\log n} \leq \eta_1/n^3$. We now define the following sets:

$$\mathcal{P}_1 \triangleq \{i \in [n] : \mathsf{U}_{it_2} \leq \eta_{3\log n}\}, \tag{101}$$

$$\mathcal{P}_2 \triangleq \bigcup_q \mathcal{X}'_q. \tag{102}$$

For each $i \in \mathcal{P}_1$, we estimate $\widehat{\mathsf{U}}_{it_2} = 0$ and since $\mathsf{U}_{it_2} \leq \mathsf{U}_{\pi_{\widehat{\boldsymbol{\mu}}_{t_1}}(n)t_2}/n^3$ and $\mathsf{U}_{\pi_{\widehat{\boldsymbol{\mu}}_{t_1}}(n)t_2} = \max_{i \in [n]} \mathsf{U}_{it_2}$, we must have

$$\left| \widehat{\mathsf{U}}_{it_2} - \mathsf{U}_{it_2} \right| \leq \frac{\max_{i \in [n]} \mathsf{U}_{it_2}}{n^3} \quad \text{for all } i \in \mathcal{P}_1. \tag{103}$$

For each $i \in \mathcal{P}_2$, we query U_{it_2} and estimate $\widehat{\mathsf{U}}_{it_2} = \mathsf{U}_{it_2}$. Notice that we have

$$[n] \setminus \{\mathcal{P}_1 \cup \mathcal{P}_2\} = \bigcup_q \mathcal{X}_q, \tag{104}$$

and therefore for each $\mathcal{X}_q$ such that $\mathcal{X}_q \neq \phi$, we estimate $\widehat{\mathsf{U}}_{it_2} = \mathsf{U}_{\pi_{\widehat{\boldsymbol{\mu}}_{t_1}}(p_{\eta_q})t_2}$ for all $i \in \mathcal{X}_q$. Now, since

$$\mathsf{U}_{\pi_{\widehat{\boldsymbol{\mu}}_{t_1}}(p_{\eta_{q-1}}-1)t_2} \leq \eta_{q-1} \quad \text{and} \quad \mathsf{U}_{\pi_{\widehat{\boldsymbol{\mu}}_{t_1}}(p_{\eta_q})t_2} \geq \eta_q = \frac{\eta_{q-1}}{2}, \tag{105}$$

we must have that for all $i \in \mathcal{X}_q$ such that $\mathcal{X}_q \neq \phi$,

$$\max_{i \in \mathcal{X}_q} \mathsf{U}_{it_2} \leq 2 \min_{i \in \mathcal{X}_q} \mathsf{U}_{it_2}, \tag{106}$$

which proves the lemma. Note that each binary search step in Procedure 8 requires $O(\log n)$ queries, and may require an additional $O(\log n)$ queries if $\mathcal{X}_q' \neq \phi$. Similarly the time-complexity of each binary step is $O(\log n)$ as well. Since we are making at most $3 \log n$ binary search steps, we get the desired query and time-complexity results. $\square$

Lemma 11 implies the following corollary.

Corollary 8. *Consider the setting of Lemma 11. Then,*

$$\sum_{i \in [n]} \left| \widehat{\mathsf{U}}_{it_2} - \mathsf{U}_{it_2} \right| \leq \sum_{i \in [n]} \mathsf{U}_{it_2} + \frac{\max_{i \in [n]} \mathsf{U}_{it_2}}{n^2}. \tag{107}$$

Proof. Recall that $\mathcal{V} \triangleq \{1, 2, \ldots, 3 \log n\}$. Then, note that

$$\sum_{i \in [n]} \left| \widehat{\mathsf{U}}_{it_2} - \mathsf{U}_{it_2} \right| = \sum_{i \in \mathcal{P}_1} \left| \widehat{\mathsf{U}}_{it_2} - \mathsf{U}_{it_2} \right| + \sum_{i \in \mathcal{P}_2} \left| \widehat{\mathsf{U}}_{it_2} - \mathsf{U}_{it_2} \right| + \sum_{q \in \mathcal{V}} \sum_{i \in \mathcal{X}_q : |\mathcal{X}_q| \neq \phi} \left| \widehat{\mathsf{U}}_{it_2} - \mathsf{U}_{it_2} \right|. \tag{108}$$

Next, we bound each of the terms on the r.h.s. of the above inequality. We have,

$$\sum_{i \in \mathcal{P}_2} \left| \widehat{\mathsf{U}}_{it_2} - \mathsf{U}_{it_2} \right| = 0, \tag{109}$$

$$\sum_{i \in \mathcal{P}_1} \left| \widehat{\mathsf{U}}_{it_2} - \mathsf{U}_{it_2} \right| \leq \frac{|\mathcal{P}_2| \max_{i \in [n]} \mathsf{U}_{it_2}}{n^3} \leq \frac{\max_{i \in [n]} \mathsf{U}_{it_2}}{n^2}. \tag{110}$$

Finally, for each $q \in \mathcal{V}$ such that $\mathcal{X}_q \neq \phi$, recall that $\widehat{\mathsf{U}}_{it_2} = \min_{j \in \mathcal{X}_q} \mathsf{U}_{it_2}$, for all $i \in \mathcal{X}_q$, and therefore,

$$\max_{i \in \mathcal{X}_q} \mathsf{U}_{it_2} \leq 2 \min_{i \in \mathcal{X}_q} \mathsf{U}_{it_2} \implies \left| \widehat{\mathsf{U}}_{it_2} - \mathsf{U}_{it_2} \right| \leq \frac{\max_{i \in \mathcal{X}_q} \mathsf{U}_{it_2}}{2} \leq \frac{\sum_{i \in \mathcal{X}_q} \mathsf{U}_{it_2}}{|\mathcal{X}_q|}, \quad \forall i \in \mathcal{X}_q.$$

Thus,

$$\sum_{q \in \mathcal{V}} \sum_{i \in \mathcal{X}_q} \left| \widehat{\mathsf{U}}_{it_2} - \mathsf{U}_{it_2} \right| \leq \sum_{i \in [n]} \mathsf{U}_{it_2},$$

which proves the desired result. $\square$

It is left to estimate the center of the smaller cluster among the two. This is done in Steps 7-13 of Procedure 5. Specifically, for each $q \in \mathcal{V}$, we randomly sample with replacement r elements from $\mathcal{X}_q$ where $\mathcal{X}_q \neq \phi$. We denote this sampled multi-set by $\mathcal{Y}_q$, and query U_{it_2}, for each $i \in \mathcal{Y}_q$. We also sample r elements from $\mathcal{P}_1$, and denote this sampled multi-set by $\mathcal{Q}$, and query U_{it_2} for each $i \in \mathcal{Q}$. Note that we have already queried U_{it_2} for every $i \in \mathcal{P}_2$ in Step 19 of Procedure 7. Subsequently, we propose the following estimate for the center of the smaller cluster,

$$\widehat{\boldsymbol{\mu}}_{t_2} = \frac{\sum_{q \in \mathcal{V} : \mathcal{X}_q \neq \phi} \frac{|\mathcal{X}_q|}{r} \sum_{i \in \mathcal{Y}_q} \mathsf{U}_{it_2}^{\alpha} \boldsymbol{x}_i + \sum_{i \in \mathcal{P}_2} \mathsf{U}_{it_2}^{\alpha} \boldsymbol{x}_i + \sum_{i \in \mathcal{Q}} \frac{|\mathcal{P}_1|}{r} \mathsf{U}_{it_2}^{\alpha} \boldsymbol{x}_i}{\sum_{q \in \mathcal{V} : \mathcal{X}_q \neq \phi} \frac{|\mathcal{X}_q|}{r} \sum_{i \in \mathcal{Y}_q} \mathsf{U}_{it_2}^{\alpha} + \sum_{i \in \mathcal{P}_2} \mathsf{U}_{it_2}^{\alpha} + \sum_{i \in \mathcal{Q}} \frac{|\mathcal{P}_1|}{r} \mathsf{U}_{it_2}^{\alpha}}. \tag{111}$$

The following result gives guarantees on the estimation error associated with the smaller cluster among the two.

Lemma 12. *Let $(\mathcal{X},\mathcal{P})$ be a consistent center-based clustering instance, and let $\delta\in(0,1)$. Then, with probability at least $1-\delta/2$, the estimator in (111) satisfies $\left\|\widehat{\boldsymbol{\mu}}_{t_2}-\boldsymbol{\mu}_{t_2}\right\|_2\leq\epsilon$, if $r\geq\frac{c'\mathsf{R}^4}{\epsilon^4}\log\frac{2}{\eta\delta}$, where $c'>0$ is an absolute constant. Also, this estimate requires $O\left(r\log n\right)$ membership-oracle queries, and a time-complexity of $O\left(dr\log n\right)$.*

Proof of Lemma 12. First, note that $[n]=\cup_{q:\mathcal{X}_q\neq\phi}\mathcal{X}_q\cup\mathcal{P}_1\cup\mathcal{P}_2$. Therefore,

$$\boldsymbol{\mu}_{t_2}=\frac{\sum_{q:\mathcal{X}_q\neq\phi}\boldsymbol{\lambda}_q+\sum_{i\in\mathcal{P}_2}\mathsf{U}_{it_2}^{\alpha}\boldsymbol{x}_i+\boldsymbol{\rho}}{\sum_{q:\mathcal{X}_q\neq\phi}\mathsf{Y}_q+\sum_{i\in\mathcal{P}_2}\mathsf{U}_{it_2}^{\alpha}+\mathsf{B}},\tag{112}$$

where $\boldsymbol{\lambda}_q\triangleq\sum_{i\in\mathcal{X}_q}\mathsf{U}_{it_2}^{\alpha}\boldsymbol{x}_i$, $\boldsymbol{\rho}\triangleq\sum_{i\in\mathcal{P}_1}\mathsf{U}_{it_2}^{\alpha}\boldsymbol{x}_i$, $\mathsf{Y}_q\triangleq\sum_{i\in\mathcal{X}_q}\mathsf{U}_{it_2}^{\alpha}$, and $\mathsf{B}\triangleq\sum_{i\in\mathcal{P}_1}\mathsf{U}_{it_2}^{\alpha}$. Similarly, using (111), we have

$$\widehat{\boldsymbol{\mu}}_{t_2}=\frac{\sum_{q\in\mathcal{V}:\mathcal{X}_q\neq\phi}\bar{\boldsymbol{\lambda}}_q+\sum_{i\in\mathcal{P}_2}\mathsf{U}_{it_2}^{\alpha}\boldsymbol{x}_i+\bar{\boldsymbol{\rho}}}{\sum_{q\in\mathcal{V}:\mathcal{X}_q\neq\phi}\bar{\mathsf{Y}}_q+\sum_{i\in\mathcal{P}_2}\mathsf{U}_{it_2}^{\alpha}+\bar{\mathsf{B}}},\tag{113}$$

where $\bar{\boldsymbol{\lambda}}_q\triangleq\frac{|\mathcal{X}_q|}{r}\sum_{i\in\mathcal{Y}_q}\mathsf{U}_{it_2}^{\alpha}\boldsymbol{x}_i$, $\bar{\boldsymbol{\rho}}\triangleq\sum_{i\in\mathcal{Q}}\frac{|\mathcal{P}_1|}{r}\mathsf{U}_{it_2}^{\alpha}\boldsymbol{x}_i$, $\bar{\mathsf{Y}}_q\triangleq\frac{|\mathcal{X}_q|}{r}\sum_{i\in\mathcal{Y}_q}\mathsf{U}_{it_2}^{\alpha}$, and $\bar{\mathsf{B}}\triangleq\sum_{i\in\mathcal{Q}}\frac{|\mathcal{P}_1|}{r}\mathsf{U}_{it_2}^{\alpha}$. Notice that for each $q\in\mathcal{V}$, the random variable $\bar{\boldsymbol{\lambda}}_q$ can be written as a sum of r i.i.d. random variables $\bar{\boldsymbol{\lambda}}_{q,i_p}\triangleq|\mathcal{X}_q|\,\mathsf{U}_{i_pt_2}^{\alpha}\boldsymbol{x}_{i_p}$, where i_p is sampled uniformly over $[n]$. Similarly, $\bar{\mathsf{Y}}_q$ written as a sum of r i.i.d random variables $\bar{\mathsf{Y}}_{q,i_p}\triangleq|\mathcal{X}_q|\,\mathsf{U}_{i_pt_2}^{\alpha}$, where again i_p is sampled uniformly over $[n]$. Finally, both $\bar{\boldsymbol{\rho}}$ and $\bar{\mathsf{B}}$ can also be written as a sum of r i.i.d random variables $\bar{\boldsymbol{\rho}}_{i_p}\triangleq|\mathcal{P}_1|\,\mathsf{U}_{i_pt_2}^{\alpha}\boldsymbol{x}_{i_p}$ and $\bar{\mathsf{B}}_{i_p}\triangleq|\mathcal{P}_1|\,\mathsf{U}_{i_pt_2}^{\alpha}$, respectively, where i_p is sampled uniformly over $[n]$. Thus, it is evident that $\mathbb{E}\bar{\boldsymbol{\lambda}}_{q,i_p}=\sum_{i\in\mathcal{X}_q}\mathsf{U}_{it_2}^{\alpha}\boldsymbol{x}_i$ and $\mathbb{E}\bar{\mathsf{Y}}_{q,i_p}=\sum_{i\in\mathcal{X}_q}\mathsf{U}_{it_2}^{\alpha}$, for all $p\in[r]$. Similarly, $\mathbb{E}\bar{\boldsymbol{\rho}}_{i_p}=\sum_{i\in\mathcal{P}_1}\mathsf{U}_{it_2}^{\alpha}\boldsymbol{x}_i$ and $\mathbb{E}\bar{\mathsf{B}}_{i_p}=\sum_{i\in\mathcal{P}_1}\mathsf{U}_{it_2}^{\alpha}$, for all $p\in[r]$. Next, we note that

$$\begin{aligned}\widehat{\boldsymbol{\mu}}_{t_2}=\boldsymbol{\mu}_{t_2}&+\sum_{q\in\mathcal{V}}\frac{\bar{\boldsymbol{\lambda}}_q-\boldsymbol{\lambda}_q}{\sum_{q\in\mathcal{V}:\mathcal{X}_q\neq\phi}\mathsf{Y}_q+\sum_{i\in\mathcal{P}_2}\mathsf{U}_{it_2}^{\alpha}+\mathsf{B}}+\frac{\bar{\boldsymbol{\rho}}-\boldsymbol{\rho}}{\sum_{q\in\mathcal{V}:\mathcal{X}_q\neq\phi}\mathsf{Y}_q+\sum_{i\in\mathcal{P}_2}\mathsf{U}_{it_2}^{\alpha}+\mathsf{B}}\\&+\widehat{\boldsymbol{\mu}}_{t_2}\cdot\frac{\sum_{q\in\mathcal{V}:\mathcal{X}_q\neq\phi}\mathsf{Y}_q-\sum_{q\in\mathcal{V}:\mathcal{X}_q\neq\phi}\bar{\mathsf{Y}}_q+\mathsf{B}-\bar{\mathsf{B}}}{\sum_{q\in\mathcal{V}:\mathcal{X}_q\neq\phi}\mathsf{Y}_q+\sum_{i\in\mathcal{P}_2}\mathsf{U}_{it_2}^{\alpha}+\mathsf{B}}.\end{aligned}\tag{114}$$

Therefore, using the triangle inequality and the fact that $\left\|\widehat{\boldsymbol{\mu}}_{t_2}\right\|_2\leq\mathsf{R}$, we get

$$\begin{aligned}\left\|\widehat{\boldsymbol{\mu}}_{t_2}-\boldsymbol{\mu}_{t_2}\right\|_2\leq&\sum_{q\in\mathcal{V}}\left\|\frac{\bar{\boldsymbol{\lambda}}_q-\boldsymbol{\lambda}_q}{\sum_{q\in\mathcal{Q}:\mathcal{X}_q\neq\phi}\mathsf{Y}_q+\sum_{i\in\mathcal{P}_2}\mathsf{U}_{it_2}^{\alpha}+\mathsf{B}}\right\|_2+\left\|\frac{\bar{\boldsymbol{\rho}}-\boldsymbol{\rho}}{\sum_{q\in\mathcal{V}:\mathcal{X}_q\neq\phi}\mathsf{Y}_q+\sum_{i\in\mathcal{P}_2}\mathsf{U}_{it_2}^{\alpha}+\mathsf{B}}\right\|_2\\&+\mathsf{R}\sum_{q\in\mathcal{V}}\left\|\frac{\bar{\mathsf{Y}}_q-\bar{Y}_q}{\sum_{q\in\mathcal{V}:\mathcal{X}_q\neq\phi}\mathsf{Y}_q+\sum_{i\in\mathcal{P}_2}\mathsf{U}_{it_2}^{\alpha}+\mathsf{B}}\right\|_2+\mathsf{R}\left\|\frac{\bar{\mathsf{B}}-\mathsf{B}}{\sum_{q\in\mathcal{V}:\mathcal{X}_q\neq\phi}\mathsf{Y}_q+\sum_{i\in\mathcal{P}_2}\mathsf{U}_{it_2}^{\alpha}+\mathsf{B}}\right\|_2.\end{aligned}\tag{115}$$

Now, for any $i\in\mathcal{Y}_q$, we have by definition,

$$\mathsf{U}_{it_2}^{\alpha}\leq\left(2\mathsf{U}_{\pi_{\widehat{\boldsymbol{\mu}}_{t_1}}(p_{\eta_q})t_2}\right)^{\alpha},\tag{116}$$

and for any $i\in\mathcal{P}_1$,

$$\mathsf{U}_{it_2}^{\alpha}\leq\frac{\max_{j\in[n]}\mathsf{U}_{jt_2}^{\alpha}}{n^3}.\tag{117}$$

Next, using Lemmas 3 and 4, we have for all $q\in\mathcal{V}$, with probability at least $1-\delta$,

$$\left\|\bar{\boldsymbol{\lambda}}_q-\boldsymbol{\lambda}_q\right\|_2\leq\mathsf{R}|\mathcal{X}_q|(2\mathsf{U}_{\pi_{\widehat{\boldsymbol{\mu}}_{t_1}}(p_{\eta_q})t_2})^{\alpha}\left(\frac{c}{r}\log\frac{2}{\eta\delta}\right)^{1/4},\tag{118}$$

$$\left|\bar{\mathsf{Y}}_q-\mathsf{Y}_q\right|\leq|\mathcal{X}_q|(2\mathsf{U}_{\pi_{\widehat{\boldsymbol{\mu}}_{t_1}}(p_{\eta_q})t_2})^{\alpha}\sqrt{\frac{c}{2r}\log\frac{2}{\eta\delta}},\tag{119}$$

$$\left\|\bar{\boldsymbol{\rho}}-\boldsymbol{\rho}\right\|_2\leq\frac{\mathsf{R}|\mathcal{P}_1|\max_{j\in[n]}\mathsf{U}_{jt_2}^{\alpha}}{n^3}\left(\frac{c}{r}\log\frac{2}{\eta\delta}\right)^{1/4},\tag{120}$$

$$\left|\bar{\mathsf{B}}-\mathsf{B}\right|\leq\frac{|\mathcal{P}_1|\max_{j\in[n]}\mathsf{U}_{jt_2}^{\alpha}}{n^3}\sqrt{\frac{c}{r}\log\frac{2}{\eta\delta}},\tag{121}$$

for some $c > 0$. Substituting the above results in (115), we get

$$\left\|\widehat{\boldsymbol{\mu}}_{t_2} - \boldsymbol{\mu}_{t_2}\right\|_2 \leq 2\mathsf{R}\left(\frac{c}{r}\log\frac{2}{\eta\delta}\right)^{1/4} \cdot \frac{\sum_{q\in\mathcal{V}:\mathcal{X}_q\neq\phi}|\mathcal{X}_q|(2\mathsf{U}_{\pi_{\widehat{\boldsymbol{\mu}}_{t_1}}(p_{\eta_q})t_1})^\alpha}{\sum_{q\in\mathcal{V}:\mathcal{X}_q\neq\phi}\mathsf{Y}_q + \sum_{i\in\mathcal{P}_2}\mathsf{U}^\alpha_{it_2} + \mathsf{B}}$$
$$+ 2\mathsf{R}\left(\frac{c}{r}\log\frac{2}{\eta\delta}\right)^{1/4} \cdot \frac{\frac{|\mathcal{P}_1|\max_{j\in[n]}\mathsf{U}^\alpha_{jt_1}}{n^3}}{\sum_{q\in\mathcal{V}:\mathcal{X}_q\neq\phi}\mathsf{Y}_q + \sum_{i\in\mathcal{P}_2}\mathsf{U}^\alpha_{it_2} + \mathsf{B}}. \tag{122}$$

Noting to the facts that

$$\sum_{q\in\mathcal{V}:\mathcal{X}_q\neq\phi}\mathsf{Y}_q = \sum_{i\in\mathcal{X}_q:\mathcal{X}_q\neq\phi}\mathsf{U}^\alpha_{it_2} \geq |\mathcal{X}_q|(\mathsf{U}_{\pi_{\widehat{\boldsymbol{\mu}}_{t_1}}(p_{\eta_q})t_1})^\alpha \tag{123}$$

and

$$\sum_{q\in\mathcal{V}:\mathcal{X}_q\neq\phi}\mathsf{Y}_q + \sum_{i\in\mathcal{P}_2}\mathsf{U}^\alpha_{it_2} + \mathsf{B} = \sum_{i\in[n]}\mathsf{U}^\alpha_{it_2}\sum_{i\in[n]}\mathsf{U}^\alpha_{it_1} \geq \max_{i\in[n]}\mathsf{U}^\alpha_{it_1}, \tag{124}$$

and finally that $|\mathcal{P}_1| \leq n$, we obtain

$$\left\|\widehat{\boldsymbol{\mu}}_{t_2} - \boldsymbol{\mu}_{t_2}\right\|_2 \leq 2^{\alpha+1}\mathsf{R}\left(\frac{c}{r}\log\frac{2}{\eta\delta}\right)^{1/4}. \tag{125}$$

Therefore, for any $\epsilon > 0$, with $r \geq \frac{c\mathsf{R}^4}{\epsilon^4}\log\frac{2}{\eta\delta}$, we obtain $\left\|\widehat{\boldsymbol{\mu}}_{t_2} - \boldsymbol{\mu}_{t_2}\right\|_2 \leq \epsilon$, which proves the lemma. □

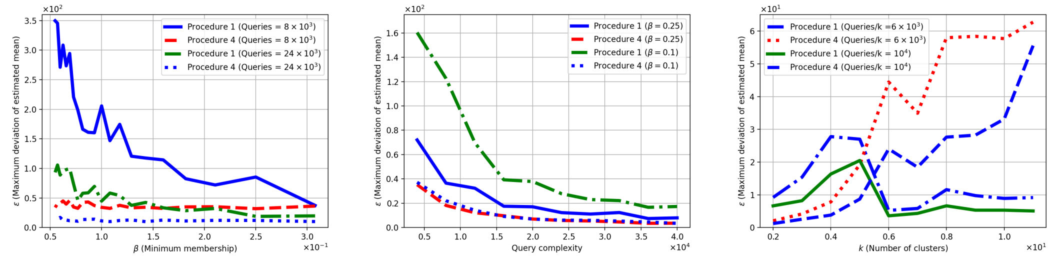
E.2 Proof of Theorem 3

First, Corollary 4 implies that a query complexity of $O\left(\frac{\mathsf{R}^4}{\epsilon^4}\log\frac{1}{\delta}\right)$, and time-complexity of $O\left(\frac{d\mathsf{R}^4}{\epsilon^4}\log\frac{1}{\delta}\right)$, suffice to approximate the center of the first cluster t_1 with probability at least $1-\delta/2$. Then, Lemma 11 allows us to estimate U_{it_2} for all $i \in [n]$ using a query complexity of $O(\log^2 n)$ and time-complexity of $O(\log^2 n)$. Also, Lemma 12 shows that a query complexity of $O\left(\frac{\mathsf{R}^4\log n}{\epsilon^4}\log\frac{1}{\eta\delta}\right)$, and time-complexity of $O\left(\frac{d\mathsf{R}^4\log n}{\epsilon^4}\log\frac{1}{\eta\delta}\right)$, suffice to approximate $\widehat{\boldsymbol{\mu}}_{t_2}$ up to an error of ϵ. Finally, we can use Lemma 7 to approximate U_{ij} up to an error of η using query complexity of $O(\log n/\eta)$, and a time-complexity of $O(n\log n + \log n/\eta)$, for all $i \in [n]$ and $j \in \{1,2\}$.

F Experiments

Synthetic Datasets: We conduct in-depth simulations of the proposed techniques over the following synthetic dataset. Specifically, we generate the dataset $\mathcal{X}$ by choosing $k = 4$ centers with dimension $d = 10$, such that $\boldsymbol{\mu}_1$ is significantly separated from the other centers; the distance from each coordinate of $\boldsymbol{\mu}_1$ to the coordinates of the other means is at least 1000. Subsequently, for each $i \in \{1,2,3,4\}$ we randomly generate L_i vectors from a spherical Gaussian distribution, with mean $\boldsymbol{\mu}_i$, and a standard deviation of 20 per coordinate. We then run the Fuzzy C-means algorithm[2], and obtain a target solution $\mathcal{P}$ to be used by the oracle for responses. In order to understand the effect of β, we fix $\mathsf{L}_1 = 5000$, and vary $\mathsf{L}_2, \mathsf{L}_3, \mathsf{L}_4 \in 5000\cdot\zeta$, where $\zeta \in \{1,2,\ldots,24\}$. It can be checked that $\beta = 4/(1+3\zeta)$. We run Algorithms 1 and 4. For the two-phase algorithm we take $\alpha = 2$, $m = \nu$, and $\eta = 0.1$, and $\alpha = 2$, $m = \nu/2.5$, $\eta_1 = 0.1$, and $\eta_2 = 0.1$, for the sequential algorithm, where $\nu \in \{2000, 6000\}$. Setting the parameters in this way keeps the same query complexity for both algorithms, so as to keep a fair comparison. We run each algorithm 20 times. For each algorithm, we evaluate the maximal error in estimating the centers. The results are shown in Fig. 1. Specifically, Fig. 1a presents the estimation error as a function of β. It can be seen that for small β's, the sequential algorithm is significantly better compared to the two-phase algorithm, whereas for larger β's, they are comparable. Then, for $\beta = 0.25, 0.1$, Fig. 1b shows the estimation error as a function of the number of queries. Finally, to understand the effect of the number of clusters, we generate k clusters

[2]https://github.com/omadson/fuzzy-c-means

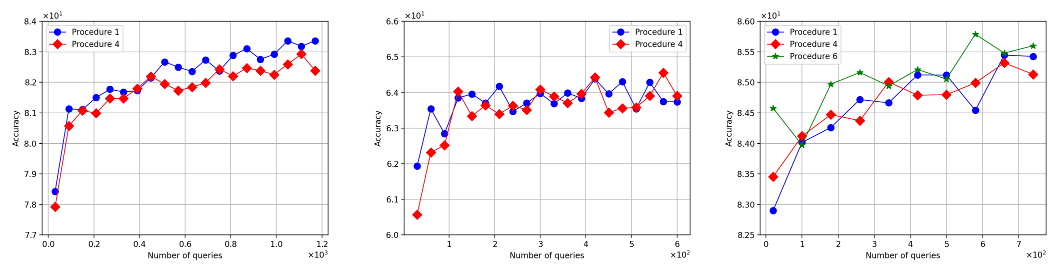


(a) Comparison of two-phase algorithm and the sequential algorithm (see Algorithm 1 and 4. The error in recovery of means is plotted with varying β.

(b) The error in recovery of means using the two-phase algorithm and sequential algorithm (see Algorithm 1 and 4 with increasing queries keeping β fixed.

(c) Comparison of the two-phase algorithm and the sequential algorithm (see Algorithm 1 and 4. The error in recovery of means is plotted with varying number of clusters (k).

Figure 1: Testing algorithms over synthetic datasets.



(a) Iris

(b) Wine

(c) Breast Cancer

Figure 2: Classification accuracy of algorithms for the Iris, Wine and Breast Cancer datasets.

using a similar method as above. We take $\mathsf{L}_1 = 1000$, and $\mathsf{L}_i = 12000$, for all $2 \leq i \leq k$. We vary $k \in \{2, 3, \ldots, 11\}$. For the two-phase algorithm, we take $\alpha = 2$, $m = \nu$, and $\eta = 0.1$, and $\alpha = 2$, $m = \nu/2.5$, $\eta_1 = 0.1$, and $\eta_2 = 0.1$, for the sequential algorithm, where $\nu = \{2000, 6000\}$. Fig. 1c shows the estimation error as a function of k. We can clearly observe that the two-phase algorithm performs significantly better as k increases but the sequential algorithm works better for small k.

Real-World Datasets: In our experiments, we use three well-known real-world datasets available in scikit-learn [41]: the Iris dataset (150 elements, 4 features, and 3 classes), the Wine dataset (178 elements, 13 features, and 3 classes), and the Breast Cancer dataset (569 elements, 30 features, and 2 classes). For the Iris and Wine datasets, we run the two-phase and sequential algorithms. We take $\alpha = 2$, $m = \nu$, and $\eta = 0.1$, for the two-phase algorithm, and $\alpha = 2$, $m = 2\nu/3$, $r = m/\eta_1$, $\eta_1 = 0.1$, and $\eta_2 = 0.1$, for the sequential algorithm, where $\nu \in \{10, 20, \ldots, 410\}$, keeping the same query complexity for both algorithms. These values do not necessarily satisfy what is needed by our theoretical results. We run both algorithms with each set of parameters 500 times to account for the randomness. In our experiments, we use a hard cluster assignment as ground truth (or rather the target clustering $\mathcal{P}$ to be used by the oracle for responses), and use our algorithms to return a fuzzy assignment. We must point out over here that our fuzzy algorithms can be used to solve hard clustering problems as well and therefore, it is not unreasonable to have hard clusters as the target solution.

Subsequently, we estimate the membership weights for all elements, and for each element, we predict the class the element belongs to as the one to which the element has the highest membership weight (i.e., $\mathsf{argmax}_j \widehat{\mathsf{U}}_{ij}$, for element i)[3]. Once we have classified all the data-points using our algorithms, we can check the classification accuracy since we possess the ground-truth labels. Note that the ground truth labels can be inconsistent with the best clustering or $\mathcal{P}^\star$, the solution that minimizes the

[3] This is similar to rounding in Linear Programming

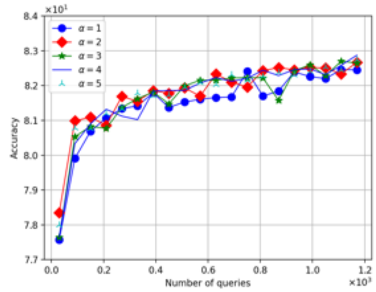


Figure 3: Classification accuracy as a function of the fuzzifier α and the number of queries for the Iris dataset.

objective in the Fuzzy k-means problems (Defintion 1) but we assume that the "ground truth" labels that are given by humans are a good proxy for the best clustering.

We then plot the classification accuracy as a function of the number of queries. Fig. 2 shows the average classification accuracy for the above three data-sets by comparing the predicted classes and the ground truth. For the Breast Cancer dataset, since the number of clusters is two, we additionally compare the two-phase and sequential algorithms to Algorithm 5 with $\alpha = 2$, $m = 2\nu/3$, $r = m/\eta_1$, and $\eta = 0.1$. It turns out that for these real-world datasets, the performance of all algorithms are comparable. It can be seen that the accuracy increases as a function of the number of queries, as expected. **Further, by using the well-known Lloyd's style iterative Fuzzy C-means algorithm with random initialization [21], we get an average classification accuracy (over 20 trials) of only** 31.33%**,** 35.96% **and** 14.58% **on the Iris, Wine, and Breast Cancer datasets, respectively**. This experiment shows that using a few membership queries increases the accuracy of a poly-time algorithm drastically, corroborating the results of our paper.

Accuracy as a function of α**:** As discussed in right after Theorem 2, the "fuzzifier" α is not subject to an optimization. Nonetheless, if we assume the existence of a ground truth, we can compare the clustering accuracy for different values of α. Accordingly, in Fig. 3 we test the performance of our algorithms on the Iris dataset for a few values of α. We calculate the average accuracy over 500 trials for each set of parameters. We conclude this section by discussing the issue of comparing our semi-supervised fuzzy approach to the semi-supervised hard objective [4]. Generally speaking, in the absence of a ground truth, comparing both approaches is *meaningless*. When the ground truth represents a disjoint clustering, then it is reasonable that following a hard approach (essentially $\alpha = 1$) will capture this ground truth better. However, the whole point of using fuzzy clustering in the first place is when the clusters, in some sense, overlap. Indeed, the initial main motivation for studying fuzzy clustering is that it is applicable to datasets where datapoints show affinity to multiple labels, the clustering criteria are vague and data features are unavailable. Nonetheless, in Fig. 3 we compare the performance of both the fuzzy and hard approaches (essentially $\alpha = 1$) on the Iris dataset.

G Discussion on noisy oracle responses

In this section, we briefly discuss the effect of a noisy membership-oracle, defined as follows.

Definition 5 (Noisy Membership-Oracle)**.** *A fuzzy query asks the membership weight of an instance* $\boldsymbol{x}_i$ *to a cluster* j *and obtains in response a noisy answer* $\mathcal{O}_{\text{noisy}}(i,j) = \mathsf{U}_{ij} + \zeta_{ij}$ *where* ζ_{ij} *is a zero mean random variable with variance* σ^2.

To present our main result, let $\rho \in \mathbb{R}_+$ be defined as

$$\min_{j\in[k]} \sum_{i\in[n]} \mathsf{U}_{ij}^{\alpha} = \rho n. \tag{126}$$

The result below handles the situation where the oracle responses are noisy.

Theorem 9. *G Let $\kappa > 0$, and assume that there exists an $(\epsilon_1, \epsilon_2, \mathsf{Q})$-solver for a clustering instance $(\mathcal{X}, \mathcal{P})$ using the membership-oracle responses O_{fuzzy}. Then, there exist a*

$$\left(\frac{2\mathsf{R}\alpha(\epsilon_2+\kappa)}{\rho}, \epsilon_2+\kappa, \frac{8\mathsf{Q}\sigma^2\log n}{\kappa^2}\right) - \mathsf{solver},$$

for $(\mathcal{X}, \mathcal{P})$ using the noisy oracle O_{noisy}.

Proof. Assume that algorithm $\mathcal{A}_{\mathsf{noiseless}}$ is an $(\epsilon_1, \epsilon_2, \mathsf{Q})$-solver for a clustering instance $(\mathcal{X}, \mathcal{P})$ using queries to a noiseless oracle O_{fuzzy}. In order to handle noisy responses we propose the following algorithm $\mathcal{A}_{\mathsf{noisy}}$: apply algorithm $\mathcal{A}_{\mathsf{noiseless}}$ for T steps using noisy queries to O_{noisy}. We will show that this algorithm obtains the guarantees in Theorem . To that end, in each such step, we obtain noisy estimates for the memberships and the centers. Then, we use these local estimates to obtain a clean final estimate for the memberships and the centers. Specifically, consider T independent and noisy clustering instances $\{\mathcal{P}^t \equiv (\boldsymbol{\mu}^t, \mathsf{V}^t)\}_{t=1}^{\mathsf{T}}$, such that

$$\mathsf{V}_{ij}^t = \mathsf{U}_{ij} + \zeta_{ij} \quad \text{and} \quad \boldsymbol{\mu}_j^t = \frac{\sum_{i=1}^n \mathsf{V}_{ij}^\alpha \boldsymbol{x}_i}{\sum_{i=1}^n \mathsf{V}_{ij}^\alpha}, \tag{127}$$

for $t \in [\mathsf{T}]$. Note that the randomness in the definition of the aforementioned clustering instances lies in the realization of the independent random variables $\zeta_{ij}^1, \zeta_{ij}^2, \ldots, \zeta_{ij}^{\mathsf{T}}$. For each such instance we apply one of the algorithms we developed for the noiseless oracle. Accordingly, for all $t \in [\mathsf{T}]$, suppose we have a $(\epsilon_1, \epsilon_2, Q)$-solver that makes Q queries to the $\mathcal{P}^t$-oracle to compute $\widehat{\mathsf{V}}_{ij}^t$. Then, we know that,

$$\left|\widehat{\mathsf{V}}_{ij}^t - \mathsf{V}_{ij}^t\right| \le \epsilon_2, \tag{128}$$

for all $t < [\mathsf{T}]$. Now, all we have to do is to use these local estimates to calculate our final estimates for the underlying memberships and centers. Specifically, for $\mathsf{T}' < \mathsf{T}$, we must have

$$\left|\frac{1}{\mathsf{T}'}\sum_{j=1}^{\mathsf{T}'} \widehat{\mathsf{V}}_{ij}^t - \mathsf{U}_{ij}\right| \le \epsilon_2 + \left|\frac{1}{\mathsf{T}'}\sum_{t=1}^{\mathsf{T}'} \zeta_{ij}^t\right|. \tag{129}$$

By Chebychev's inequality, for any $\kappa > 0$, we get

$$\Pr\left[\left|\frac{1}{\mathsf{T}'}\sum_{t=1}^{\mathsf{T}'} \zeta_{ij}^t\right| \ge \kappa\right] \le \frac{\sigma^2}{\kappa^2 \mathsf{T}'}. \tag{130}$$

Next, we partition the T responses from the oracle into B batches of size T' each. For batch $b \in [\mathsf{B}]$, define the random variable $\mathsf{Y}^b \triangleq \mathbb{1}\left[\left|\frac{1}{\mathsf{T}'}\sum_{t \in \text{Batch } b} \zeta_{ij}^t\right| \ge \kappa\right]$. Clearly, $\Pr(\mathsf{Y}^b = 1) \le \frac{\sigma^2}{\kappa^2 \mathsf{T}'}$ and further, $\mathsf{Y}^1, \mathsf{Y}^2, \ldots, \mathsf{Y}^{\mathsf{B}}$ are independent random variables. Therefore, Chernoff bound implies that

$$\Pr\left(\sum_{b=1}^{\mathsf{B}} \mathsf{Y}^b \ge \mathsf{B}/2\right) \le \exp\left[-2\mathsf{B}\left(\frac{1}{2} - \frac{\sigma^2}{\kappa^2 \mathsf{T}'}\right)^2\right]. \tag{131}$$

Our final membership estimate is evaluated as follows:

$$\widehat{\mathsf{U}}_{ij} \triangleq \mathsf{median}\left(\frac{1}{\mathsf{T}'}\sum_{t \in \text{Batch } 1} \mathsf{V}_{ij}^t, \frac{1}{\mathsf{T}'}\sum_{t \in \text{Batch } 2} \mathsf{V}_{ij}^t, \ldots, \frac{1}{\mathsf{T}'}\sum_{t \in \text{Batch } \mathsf{B}} \mathsf{V}_{ij}^t\right), \tag{132}$$

namely, $\widehat{\mathsf{U}}_{ij}$ is the median of the mean of $\widehat{\mathsf{V}}_{ij}^t$ in each batch. Therefore, for $\mathsf{B} = 6\log n$ and $\mathsf{T}' = 4\sigma^2/\kappa^2$ (hence $\mathsf{T} = 8\sigma^2 \log n/\kappa^2$), we must have that

$$\Pr\left(\left|\widehat{\mathsf{U}}_{ij} - \mathsf{U}_{ij}\right| \ge \epsilon_2 + \kappa\right) \le \frac{2}{n^3}. \tag{133}$$

Therefore, by taking a union bound over all $i \in [n], j \in [k]$, we can compute $\widehat{\mathsf{U}}_{ij}$, an estimate of U_{ij}, such that

$$\left|\widehat{\mathsf{U}}_{ij} - \mathsf{U}_{ij}\right| \le \epsilon_2 + \kappa, \tag{134}$$

for all $i \in [n], j \in [k]$ with probability at least $1 - 1/n$. Finally, we estimate the means $\boldsymbol{\mu}_j$'s using the already computed $\widehat{\mathsf{U}}_{ij}$'s as follows. Note that,

$$\widehat{\boldsymbol{\mu}}_j = \frac{\sum_i \widehat{\mathsf{U}}_{ij}^{\alpha} \boldsymbol{x}_i}{\sum_i \widehat{\mathsf{U}}_{ij}^{\alpha}} \triangleq \frac{\widehat{\boldsymbol{\lambda}}_{\boldsymbol{x}}}{\widehat{\mathsf{Y}}}, \tag{135}$$

and

$$\boldsymbol{\mu}_j = \frac{\sum_{i \in [n]} \mathsf{U}_{ij}^{\alpha} \boldsymbol{x}_i}{\sum_{i \in [n]} \mathsf{U}_{ij}^{\alpha}} \triangleq \frac{\boldsymbol{\lambda}_{\boldsymbol{x}}}{\mathsf{Y}}. \tag{136}$$

Therefore, we get

$$\left\|\widehat{\boldsymbol{\mu}}_j - \boldsymbol{\mu}_j\right\|_2 \leq \left\|\frac{\widehat{\boldsymbol{\lambda}}_{\boldsymbol{x}} - \boldsymbol{\lambda}_{\boldsymbol{x}}}{\mathsf{Y}}\right\|_2 + \left\|\frac{\widehat{\boldsymbol{\lambda}}_{\boldsymbol{x}}}{\widehat{\mathsf{Y}}} \frac{\mathsf{Y} - \widehat{\mathsf{Y}}}{\mathsf{Y}}\right\|_2 \leq \left\|\frac{\widehat{\boldsymbol{\lambda}}_{\boldsymbol{x}} - \boldsymbol{\lambda}_{\boldsymbol{x}}}{\mathsf{Y}}\right\|_2 + \mathsf{R}\left|\frac{\widehat{\mathsf{Y}} - \mathsf{Y}}{\mathsf{Y}}\right|. \tag{137}$$

Using (134) it is evident that

$$\left|\widehat{\mathsf{Y}} - \mathsf{Y}\right| \leq \alpha n(\epsilon_2 + \kappa)(1 + o(1)), \tag{138}$$

$$\left\|\widehat{\boldsymbol{\lambda}}_{\boldsymbol{x}} - \boldsymbol{\lambda}_{\boldsymbol{x}}\right\|_2 \leq \mathsf{R}\alpha n(\epsilon_2 + \kappa)(1 + o(1)). \tag{139}$$

Combining (137)–(139) together with (126), we finally obtain that

$$\left\|\widehat{\boldsymbol{\mu}}_j - \boldsymbol{\mu}_j\right\|_2 \leq \frac{2\mathsf{R}\alpha(\epsilon_2 + \kappa)}{\rho}, \tag{140}$$

for all $j \in [k]$, which concludes the proof. □

H Membership queries from similarity queries

Recall that $\mathcal{X} \subset \mathbb{R}^d, |\mathcal{X}| = n$ is the set of points provided as input along with their corresponding d-dimensional vector assignments denoted by $\{\boldsymbol{x}_i\}_{i=1}^n$. Recall that the membership-oracle $\mathcal{O}_{\text{fuzzy}}(i,j) = \mathsf{U}_{ij}$ returns the membership weight of the instance $\boldsymbol{x}_i$ to a cluster j. However, such oracle queries are often impractical in real-world settings since it requires knowledge of the relevant clusters. Instead, a popular query model that takes a few elements (two or three) as input and is easy to implement in practice is the following similarity query "How similar are these elements?" [4] showed that for the hard clustering setting, a membership query can be simulated by k pairwise similarity queries since a pairwise similarity query reveals whether two items belong to the same cluster or not in the hard clustering setting. In the fuzzy problem we model the oracle response to the similarity query by the inner product of their membership weight vectors. More formally, we have

Definition 6 (Restatement of Definition 3). *A fuzzy pairwise similarity query asks the similarity of two distinct instances $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ i.e., $\mathcal{O}_{\text{sim}}(i,j) = \langle \mathsf{U}_i, \mathsf{U}_j \rangle$. A fuzzy triplet similarity query asks the similarity of three distinct instances $\boldsymbol{x}_p, \boldsymbol{x}_q, \boldsymbol{x}_r$ i.e. $\mathcal{O}_{\text{triplet}}(p,q,r) = \sum_{t \in [k]} \mathsf{U}_{pt}\mathsf{U}_{qt}\mathsf{U}_{rt}$.*

Now, we show that fuzzy pairwise similarity queries can often be used to simulate $\mathcal{O}_{\mathsf{fuzzy}}(i,j)$. Note that if we possess the membership weight vectors of k elements that are linearly independent, then, for a new element, responses to fuzzy pairwise similarity queries with the aforementioned k elements reveals all the membership weights of the new element. Now, the question becomes "How can we obtain the membership weights of the k elements in the first place?". Suppose we sub-sample a set of elements $\mathcal{Y} \subseteq \mathcal{X}$ such that $|\mathcal{Y}| = m > k$ and we make all fuzzy pairwise similarity queries among the elements present in $\mathcal{Y}$. Let us denote by V the membership weight matrix U constrained to the rows corresponding to the elements in $\mathcal{Y}$. Clearly, the fuzzy pairwise similarity queries between all pairs of elements in $\mathcal{Y}$ reveals $\mathsf{V}\mathsf{V}^T$, the gram matrix of V. If we can recover V uniquely from $\mathsf{V}\mathsf{V}^T$, and V is full rank, then we are done. If we assume almost any continuous distribution according to which the membership weight vectors are generated, then with probability 1, the matrix V is full rank. On the other hand, the question of uniquely recovering V from $\mathsf{V}\mathsf{V}^T$ is trickier. In general it is not possible to recover V uniquely from $\mathsf{V}\mathsf{V}^T$ since $\mathsf{V}\mathsf{R}$, for any orthonormal matrix R, also has the gram matrix $\mathsf{V}\mathsf{V}^T$. However, recall that in our case, the entries of V are non-negative and furthermore, the rows of V add up-to 1 leading to additional constraints. This leads to the problem

$$\text{Find } \mathsf{M} \text{ such that } \mathsf{M}\mathsf{M}^T = \mathsf{V}\mathsf{V}^T \text{ subject to } \mathsf{M} \in \mathbb{R}_{\geq 0}^{m \times k}, \sum_{j \in [k]} \mathsf{M}_{ij} = 1 \; \forall \, i \in [m].$$

As a matter of fact, this is a relatively well-studied problem known as the Symmetric Non-Negative matrix factorization or SNMF. We will say that the solution to the SNMF problem is unique if VP is the only solution to the problem for any permutation matrix P. Below, we state the following sufficient condition that guarantees the uniqueness of the solution to the SNMF problem.

Lemma 13 (Lemma 4 in [26])**.** *If* $\mathsf{rank}(\mathsf{V}) = k$*, then the solution to the SNMF problem is unique if and only if the non- negative orthant is the only self-dual simplicial cone* $\mathcal{A}$ *with* k *extreme rays that satisfies* $\mathsf{cone}(\mathsf{V}^T) \subseteq \mathcal{A} = \mathcal{A}^\star$ *where* $\mathcal{A}^\star$ *is the dual cone of* $\mathcal{A}$*, defined as* $\mathcal{A}^\star = \{\boldsymbol{y} \mid \boldsymbol{x}^T\boldsymbol{y} \geq 0 \; \forall \boldsymbol{x} \in \mathcal{A}\}$.

More recently, Lemma 13 was used in [34] to show the following result that is directly applicable to our setting:

Lemma 14. *If* V *contains any permutation matrix of dimensions* $k \times k$*, then the solution of the SNMF problem is unique.*

Suppose we have the guarantee that for each cluster $j \in [k]$, there exists a set $\mathcal{Z}_j$ of at least ρn elements belonging purely to the j^{th} cluster i.e. $\mathsf{U}_{ij} = 1$ for all $i \in \mathcal{Z}_j$. Then, for $m \geq \rho^{-1}\log(nk)$, the matrix V will contain a permutation matrix with probability at least $1 - n^{-1}$. As a results, this will lead to an overhead of $O(m^2) = O(\rho^{-2}\log^2(nk))$ queries.

If it is possible to make more complex similarity queries such as the fuzzy triplet similarity query, we can significantly generalize and improve the previous guarantees. Before proceeding further, let us provide some background on tensors beginning with the following definition:

Definition 7 (Kruskal rank)**.** *The Kruskal rank of a matrix* $\boldsymbol{A}$ *is defined as the maximum number* r *such that any* r *columns of* $\boldsymbol{A}$ *are linearly independent.*

Consider a tensor $\mathcal{A}$ of order $w \in \mathbb{N}$ for $w > 2$ on $\mathbb{R}^n$, denoted by $\mathcal{A} \in \mathbb{R}^n \otimes \mathbb{R}^n \otimes \cdots \otimes \mathbb{R}^n$ (w times). Let $\mathcal{A}_{i_1,i_2,\dots,i_w}$ where $i_1, i_2, \dots, i_w \in \{0, 1, \dots, n-1\}$, denote the element in $\mathcal{A}$ whose location along the j^{th} dimension is $i_j + 1$, i.e., there are i_j elements along the j^{th} dimension before $\mathcal{A}_{i_1,i_2,\dots,i_w}$. Notice that this indexing protocol uniquely determines the element within the tensor. For a detailed review of tensors, we defer the reader to [29]. In this work, we are interested in low-rank decomposition of tensors. A tensor $\mathcal{A}$ can be described as a rank-1 tensor if it can be expressed as[4]

$$\mathcal{A} = \underbrace{\boldsymbol{z} \otimes \boldsymbol{z} \otimes \cdots \otimes \boldsymbol{z}}_{w \text{ times}}$$

for some $\boldsymbol{z} \in \mathbb{R}^n$, i.e., $\mathcal{A}_{i_1,i_2,\dots,i_w} = \prod_{j=1}^{w} \boldsymbol{z}_{i_j}$. For a given tensor $\mathcal{A}$, we are concerned with the problem of uniquely decomposing $\mathcal{A}$ into a sum of R rank-1 tensors. A tensor $\mathcal{A}$ that can be expressed in this form is denoted as a rank-R tensor, and such a decomposition is also known as the Canonical Polyadic (CP) decomposition. Below, we state a result due to [45] describing the sufficient conditions for the unique CP decomposition of a rank-R tensor $\mathcal{A}$.

Lemma 15 (Unique CP decomposition [45])**.** *Suppose* $\mathcal{A}$ *is the sum of* R *rank-1 tensors, i.e.,*

$$\mathcal{A} = \sum_{r=1}^{R} \underbrace{\boldsymbol{z}^r \otimes \boldsymbol{z}^r \otimes \cdots \otimes \boldsymbol{z}^r}_{w \;\; times},$$

and further, the Kruskal rank of the $n \times R$ *matrix whose columns are formed by* $\boldsymbol{z}^1, \boldsymbol{z}^2, \dots, \boldsymbol{z}^R$ *is* J*. Then, if* $wJ \geq 2R + (w-1)$*, then the CP decomposition is unique and we can recover the vectors* $\boldsymbol{z}^1, \boldsymbol{z}^2, \dots, \boldsymbol{z}^R$ *up to permutations.*

Notice that for the special case of $w = 3$, the underlying vectors $\boldsymbol{z}^1, \boldsymbol{z}^2, \dots, \boldsymbol{z}^R$ can be recovered uniquely if they are linearly independent. Now, we are ready to show that k fuzzy triplet similarity

[4]In this work, we focus on the special case of rank-1 tensors where every component is identical. In general, rank-1 tensors can be described as $\mathcal{A} = \boldsymbol{z}^1 \otimes \boldsymbol{z}^2 \otimes \cdots \otimes \boldsymbol{z}^w$ for $\boldsymbol{z}^1, \dots \boldsymbol{z}^w \in \mathbb{R}^n$.

Algorithm 9 JENNRICH'S ALGORITHM($\mathcal{A}$)

Input: A symmetric rank-R tensor $\mathcal{A} \in \mathbb{R}^n \otimes \mathbb{R}^n \otimes \mathbb{R}^n$ of order 3.

Choose $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$ uniformly at random such that it satisfies $\|\boldsymbol{a}\|_2 = \|\boldsymbol{b}\|_2 = 1$.
Compute $\boldsymbol{T}^{(1)} \triangleq \sum_{i\in[n]} \boldsymbol{a}_i \mathcal{A}_{\cdot,\cdot,i}, \boldsymbol{T}^{(2)} \triangleq \sum_{i\in[n]} \boldsymbol{b}_i \mathcal{A}_{\cdot,\cdot,i}$.
if $\mathsf{rank}(T^1) < R$ **then**
    Return Error
end if
Solve the general eigenvalue problem $\boldsymbol{T}^{(1)}\boldsymbol{v} = \lambda_v \boldsymbol{T}^{(2)}\boldsymbol{v}$.
Return the eigen-vectors $\boldsymbol{v}$ corresponding to the non-zero eigen-values.

queries can be used to recover the memberships weights of k elements uniquely. As before, we can sub-sample a set of elements $\mathcal{Y} \subseteq \mathcal{X}$ such that $|\mathcal{Y}| = k$ and we make all possible $\binom{k}{3}$ fuzzy triplet similarity queries among the elements present in $\mathcal{Y}$. Again, let us denote by V the membership weight matrix U constrained to the rows corresponding to the elements in $\mathcal{Y}$. Let us denote by $\boldsymbol{v}^1, \boldsymbol{v}^2, \ldots, \boldsymbol{v}^k$ the k columns of the matrix V. Notice that the responses to all the fuzzy triplet similarity queries reveals the following symmetric tensor

$$\sum_{r=1}^{k} \boldsymbol{v}^r \otimes \boldsymbol{v}^r \otimes \boldsymbol{v}^r.$$

Suppose the matrix $\mathcal{V}$ is full rank. This will happen with probability 1 if the membership weights are assumed to be generated according to any continuous distributions. Algorithmically, Jennrich's algorithm (see Section 3.3, [39]) can be used to efficiently recover the unique CP decomposition of a third order low rank tensor whose underlying vectors are full rank. We have provided the algorithm (see, Algorithm 9) for the sake of completeness.

I Conclusion and outlook

In this paper, we studied the fuzzy k-means problem, and proposed a semi-supervised active clustering framework, where the learner is allowed to interact with a membership-oracle, asking for the memberships of a certain set of chosen items. We studied both the query and computational complexities of clustering in this framework. In particular, we provided two probabilistic algorithms (two-phase and sequential) for fuzzy clustering that ask $O(\mathsf{poly}(k)\log n)$ membership queries and run with polynomial-time-complexity. The main difference between these two algorithms is the dependency of their query complexities on the size of the smallest cluster β. The sequential algorithm exhibits more graceful dependency on β. Finally, for $k = 2$ we were able to remove completely the dependency on β (see, Appendix E). We hope our work has opened more doors than it closes. Apart from tightening the obtained query complexities, there are several exciting directions for future work:

- It is important to understand completely the dependency of the query complexity on β. Indeed, we showed that for $k = 2$ there exists an algorithm whose query complexity is independent of β, but what happens for $k > 2$?
- It would be interesting to understand to what extent the algorithms and analysis in this paper, can be applied to other clustering problems which depend on different metrics other than the Euclidean one.
- Our paper presents upper bounds (sufficient conditions) on the query complexity. It is interesting and challenging to derive algorithm-independent information-theoretic lower bounds on the query complexity.
- As mentioned in the introduction it is not known yet whether the fuzzy k-means problem lies in NP like the hard k-means problem. Answering this question will give a solid motivation to the semi-supervised setting considered in this paper. Furthermore, just as the information-theoretic lower bounds, it would be interesting to derive computational lower bounds as well.
- In this paper we focused on the simplest form of oracle responses. However, there are many other interesting and important settings, e.g., the noisy setting (Appendix G). Another

interesting problem would be to consider adversarial oracles who intentionally provide corrupted responses.