On Coreset Constructions for the Fuzzy $K$-Means Problem

Johannes Blömer, Sascha Brauer, Kathrin Bujna
Department of Computer Science
Paderborn University
Fürstenallee 11, 33102 Paderborn, Germany

December 23, 2016

Abstract

In this paper, we present coreset constructions for the fuzzy $K$-means problem. First, we show that one can construct a weak coresets for fuzzy $K$-means. Second, we show that there are coresets for fuzzy $K$-means with respect to balanced fuzzy $K$-means solutions. Third, we use these coresets to develop a randomized approximation algorithm whose runtime is polynomial in the number of the given points and the dimension of these points.

1 Introduction

[Dunn, 1973] was the first to present a fuzzy $K$-means objective function, which was later extended by [Bezdek et al., 1984]. Today, fuzzy $K$-means has found numerous practical applications, for example in image segmentation and biological data analysis (see [Rezaee et al., 2000] and [Dembélé and Kastner, 2003]), to name just a few.

1.1 Fuzzy $K$-Means Problem

Let $X = \{(x_1, w_1), \ldots, (x_N, w_N)\}$ be a set of data points $x_n \in \mathbb{R}^D$ weighted by $w_n \in \mathbb{R}_+$. We want to group $X$ into some predefined number of clusters $K$. These clusters are represented by mean vectors $\mu_1, \ldots, \mu_K \in \mathbb{R}^D$. In a fuzzy clustering, each data point $x_n$ belongs to each cluster, represented by a mean $\mu_k$, with a certain membership value $r_{nk} \in [0, 1]$. The fuzzy $K$-means problem has an additional parameter, the so-called fuzzifier $m \in \mathbb{N}_{\geq 2}$, which is chosen in advance and is not subject to optimization.
Definition 1 (Fuzzy K-means). Given $X = \{(x_n, w_n)\}_{n \in [N]} \subset \mathbb{R}^D \times \mathbb{R}_+$, $K \geq 1$ and $m \geq 2$, the fuzzy K-means problem is to find $C = \{\mu_k\}_{k \in [K]} \subset \mathbb{R}^D$ and $R = \{r_{nk}\}_{n \in [N], k \in [K]} \subset [0, 1]$ minimizing

$$
\phi^{(m)}_X(C, R) = \sum_{n=1}^N \sum_{k=1}^K r_{nk}^m w_n \|x_n - \mu_k\|^2_2,
$$

subject to: $\sum_{k=1}^K r_{nk} = 1$ for all $n \in [N]$.

We denote the costs of an optimal fuzzy K-means solution by $\phi^{OPT}_X(X, K, m)$. Furthermore, let $w_{\text{max}} = \max_{n \in [N]} w_n$ and $w_{\text{min}} = \min_{n \in [N]} w_n$.

In simple terms, the fuzzifier $m$ determines how much clusters are allowed to overlap, i.e., how soft the clustering is. Strictly speaking, the problem degenerates if $m$ is allowed to grow, as the cost of every solution can always be reduced by increasing $m$. Practitioners mostly choose $m = 2$. For $m = 1$, this problem would coincide with the classical K-means problem, while for $m \to \infty$ the memberships and every mean of every optimal solution converge to a uniform distribution and the center of the data set $X$, respectively. Our problem definition is a generalization of the original definition presented in [Bezdek et al., 1984] in that we consider weighted data sets. By setting all weights to 1, we obtain the original definition.

1.2 Fuzzy K-Means (FM) Algorithm

The most widely used heuristic for the fuzzy K-means problem is an alternating optimization algorithm known as fuzzy K-means (FM) algorithm. It is defined by the following first-order optimality conditions (see [Bezdek et al., 1987]):

Fixing the means $\{\mu_k\}_{k \in [K]}$, optimal memberships are given by

$$
r_{nk} = \frac{\|x_n - \mu_k\|^{-2} - \frac{2}{m-1}}{\sum_{l=1}^K \|x_n - \mu_l\|^{-2} - \frac{2}{m-1}}
$$

if $x_n \neq \mu_l$ for all $l \in [K]$. If $x_n$ coincides with some of the $\mu_l$, then the membership of $x_n$ can be distributed arbitrarily among those $\mu_l$ with $\mu_l = x_n$. Fixing the memberships $\{r_{nk}\}_{n,k}$, the optimal means are given by

$$
\mu_k = \frac{\sum_{n=1}^N r_{nk}^m w_n x_n}{\sum_{n=1}^N r_{nk}^m w_n}.
$$
1.3 Relation to the $K$-Means Problem

There is a coarse but still useful relation between the $K$-means and the fuzzy $K$-means cost function.

**Definition 2 ($K$-means).** For $X = \{(x_n, w_n)\}_{n \in [N]} \subset \mathbb{R}^D \times \mathbb{R}_+$ and $C = \{\mu_k\}_{k \in [K]} \subset \mathbb{R}^D$ we define $\text{km}_X(C) := \sum_{n=1}^{N} w_n \min_{k \in [K]} \|x_n - \mu_k\|_2^2$.

**Lemma 3.** Let $X \subset \mathbb{R}^D \times \mathbb{R}_+$, $m \in \mathbb{N}$, and $C \subset \mathbb{R}^D$ with $|C| = K$. Then, \[
\frac{1}{K^{m-1}} \text{km}_X(C) \leq \phi_X^{(m)}(C) \leq \text{km}_X(C).
\]

**Proof.** can be found in [Blömer et al., 2015].

2 Our Contribution

We present the first coreset constructions for the fuzzy $K$-means problem. Our three main contributions are the following: First, we show that one can construct a weak coresets for fuzzy $K$-means. Second, we show that there are coresets for fuzzy $K$-means with respect to balanced fuzzy $K$-means solutions. Third, we show that these coresets can be used to construct a randomized approximation algorithm whose runtime is polynomial in the number of the given points and the dimension of these points.

The following sections are organized as follows: In Section 3 we present our construction of weak coresets. Section 4 deals with coresets for balanced solutions. Finally, in Section 5 we show how to construct a randomized approximation algorithm.

3 A Weak Coreset for Fuzzy $K$-Means

The coreset construction of [Chen, 2009] can be used to obtain a $\Gamma$-Weak $(K, \epsilon)$-coreset for a fuzzy $K$-means problem, which is also a $\Gamma$-Weak $(K, \epsilon/K^{m-1})$-coreset for the $K$-means problem.

**Definition 4 ($\Gamma$-Weak $(K, \epsilon)$-Coresets).** Let $X \subset \mathbb{R}^D$, $\Gamma \subset \mathbb{R}^D$, $\epsilon \in [0, 1]$, and $K \in \mathbb{N}$. Let $S \subset X \times \mathbb{R}$ be a weighted set.

We call $S$ a $\Gamma$-weak $(K, \epsilon)$-coreset for the fuzzy $K$-means problem if, for all $C \subset \Gamma$, $|C| \leq K$, we have $(1 - \epsilon) \phi_X^{(m)}(C) \leq \phi_S^{(m)}(C) \leq (1 + \epsilon) \phi_X^{(m)}(C)$.

We call $S$ a $\Gamma$-weak $(K, \epsilon)$-coreset for the $K$-means problem if, for all $C \subset \Gamma$, $|C| \leq K$, we have $(1 - \epsilon) \text{km}_X(C) \leq \text{km}_S(C) \leq (1 + \epsilon) \text{km}_X(C)$.
Theorem 5. There is an algorithm that, given $X \subset \mathbb{R}^D$, finite set $\Gamma \subset \mathbb{R}^D$, $K \in \mathbb{N}$, $\epsilon \in [0, 1]$, and $\delta \in [0, 1]$, computes a set $S \subset X \times \mathbb{R}$ that satisfies the following property with probability at least $1 - \delta$:

For all $C \subset \Gamma$, $|C| \leq K$,

$$(1 - \epsilon) \phi^{(m)}_X(C) \leq \phi^{(m)}_S(C) \leq (1 + \epsilon) \phi^{(m)}_X(C)$$

and, additionally, for all $C \subset \Gamma$, $|C| \leq K$,

$$(1 - \epsilon/K^{m-1}) \text{km}_X(C) \leq \text{km}_S(C) \leq (1 + \epsilon/K^{m-1}) \text{km}_X(C)$$.

The set $S$ has size

$$|S| = \mathcal{O} \left( \log(N) \log \log(N) \cdot K^{2m+1} \epsilon^{-2} \log(\gamma) \log(\delta^{-1}) \right).$$

The algorithms’ runtime is bounded by

$$\mathcal{O} \left( ND \cdot K^{2m+1} \epsilon^{-2} \log(\gamma) \log(\delta^{-1}) \right).$$

Proof. can be found in Appendix B. \qed

4 A Coreset for Fuzzy $K$-Means

With the help of the results from the previous section, we can construct a coreset for balanced solutions of a fuzzy $K$-means problem, which we define as follows.

Definition 6 (Coresets for Balanced Solutions). Let $X \subset \mathbb{R}^D \times \{1\}$, $K \in \mathbb{N}$, and $\epsilon \in [0, 1]$. We call a set $S \subset X \times \mathbb{R}$ a $(K, \epsilon)$-coreset for balanced solutions with respect to $X$ if the following property is satisfied: For all $C \subset \mathbb{R}^D$, $|C| = K$, with corresponding optimal membership values $\{r_{nk}\}_{n\in[N], k\in[K]}$ with respect to $X$, where

$$R_k = \sum_{n=1}^N r_{nk}^m \geq \left( \frac{\epsilon}{4mK^2} \right)^m$$

for all $k \in [K]$, we have

$$(1 - \epsilon) \phi^{(m)}_X(C) \leq \phi^{(m)}_S(C) \leq (1 + \epsilon) \phi^{(m)}_X(C).$$

We can compute a coreset for balanced solutions as follows.
Theorem 7 (Coresets for Balanced Solutions). There is an algorithm that, given $X \subset \mathbb{R}^D \times \{1\}$, $K \in \mathbb{N}$, $\epsilon \in [0,1]$ and $\delta \in [0,1]$, computes a set $S \subset X \times \mathbb{R}$ such that, with probability at least $1 - \delta$, $S$ is a $(K, \epsilon)$-coreset for balanced solutions with respect to $X$.

The set $S$ has size

$$|S| = \mathcal{O} \left( \log(N) \log(\log(N))^2 \cdot D \cdot K^{4m} \cdot m^2 \cdot \epsilon^{-7} \cdot \log(\delta^{-1}) \right).$$

The algorithms’ runtime is bounded by $\mathcal{O}(|S| \cdot N)$.

Proof. can be found in Appendix C.

Interestingly, a coreset for balanced solutions does preserve approximative solutions in the following way.

Theorem 8 (Preserve Approximations). Let $X \subset \mathbb{R}^D \times \{1\}$, $K \in \mathbb{N}$, and $\epsilon \in [0,1]$. Let $S \subset X \times \mathbb{R}$ be a $(K, \epsilon)$-coreset for balanced solutions with respect to $X$. Then, for all $C \subset \mathbb{R}^D$, $|C| = K$, we have

$$\phi^m_S(C) \leq c \cdot \phi^{OPT}_{(S,K,m)} \Rightarrow \phi^m_X(C) \leq (1 + \epsilon) \cdot c \cdot \phi^{OPT}_{(X,K,m)}.$$

Proof. can be found in Appendix D.

5 An Application of Our Coreset for Fuzzy K-Means

In the following we use our coreset construction from the previous section to improve our results from [Blömer et al., 2015].

5.1 Results from [Blömer et al., 2015]

In [Blömer et al., 2015] we already proved the following results.

Theorem 9. There is a deterministic algorithm that, given $X = \{(x_n, w_n)\}_{n \in [N]} \subset \mathbb{R}^D \times \mathbb{R}_+$, $K \in \mathbb{N}$, $m \in \mathbb{N}$, and $\epsilon \in (0,1]$, computes a solution $(C, R)$ such that

$$\phi^m_X(C, R) \leq (1 + \epsilon) \phi^{OPT}_{(X,K,m)}.$$

The algorithms’ runtime is bounded by $N \left( \log(N) + \log \left( \frac{w_{\text{max}}}{w_{\text{min}}} \right) \right)^K K^{\mathcal{O}(mK^2D \log(1/\epsilon))}$.

Proof. can be found in [Blömer et al., 2015].
Lemma 10 (Dimension Reduction). There is a randomized algorithm that, given \( X = \{(x_n, w_n)\}_{n \in [N]} \subset \mathbb{R}^D \times \mathbb{R}_+\), \( m \in \mathbb{N}\), \( K \in \mathbb{N}\), and \( \epsilon \in [0, 1]\), computes a number \( \tilde{D} = \mathcal{O}(\log(N)/\epsilon^2) \) and a linear map \( \pi : \mathbb{R}^D \rightarrow \mathbb{R}^{\tilde{D}}\) such that with constant probability for all \( \alpha \geq 1\) it holds that memberships which induce an \( \alpha\)-approximation of the fuzzy \( K\)-means problem on \( \tilde{X} = \{\pi(x_n)\}_{n \in [N]}\) also induce a \((1 + \epsilon) \cdot \alpha\)-approximation on \( X\). The set \( \tilde{X}\) can be computed in time \( \mathcal{O}(D N \log(N)/\epsilon^2)\).

5.2 A Randomized \((1 + \epsilon)\)-Approximation Algorithm

By combining Theorem 9 with dimension reduction technique from Lemma 10 and our coreset construction for balanced fuzzy \( K\)-means solutions, we obtain a significant speedup and get rid of the exponential dependence on \( D\).

Corollary 11. There is a randomized algorithm that, given \( X \subset \mathbb{R}^D \times \{1\}\), \( K \in \mathbb{N}\), \( m \in \mathbb{N}\), and \( \epsilon \in (0, 1]\), computes a solution \((C, R)\) such that with constant probability

\[
\phi^{(m)}_X(C, R) \leq (1 + \epsilon)\phi^{OPT}_{(X,K,m)}.
\]

The algorithms’ runtime is bounded by

\[
N(\log(N))^{\mathcal{O}(mK^3/\epsilon^4)} \cdot D^{\mathcal{O}(\log(K)/\epsilon^4)} \cdot K^{\mathcal{O}((m \log(K))/\epsilon^4)}.
\]

Proof. can be found in Appendix E.

More precisely, we combine the aforementioned techniques as follows. First, we apply Theorem 4 to compute a coreset for balanced fuzzy \( K\)-means solutions. Then, we apply the dimension reduction technique from Lemma 10 to obtain a set \( \pi(S) \subset \mathbb{R}^{\tilde{D}} \times \mathbb{R}_+\) with smaller dimension \( \tilde{D} \ll D\). Finally, we apply the algorithm from Theorem 9 to the set \( \pi(S)\) and uses the result to compute the desired approximation.

Note that, since the coreset construction can only be used for unweighted data sets, the algorithm takes an unweighted data set (i.e. \( X \subset \mathbb{R}^D \times \{1\}\)) as input. However, since the reduced set \( \pi(S)\) is a weighted data set, we then need to use an approximation algorithm that can handle a weighted data set (e.g., the algorithm from Theorem 9).

In comparison to the algorithm from Theorem 9 the algorithm from Corollary 11 has the advantage that its runtime has no exponential dependence on \( D\), but the dependence on \( K\) is worse.
A Preliminaries

Definition 12 (K-Means Costs). For weighted point sets \( X = \{(x_n, w_n)\}_{n \in [N]} \subset \mathbb{R}^D \times \mathbb{R}_+ \), \( x \in \mathbb{R}^D \), and finite sets \( M \subset \mathbb{R}^D \) we let

\[
    d(x, M) := \min_{m \in M} \|x - m\|_2 ,
\]

\[
    \text{km}_X(M) := \sum_{n=1}^{N} w_n d(x_n, M)^2 \quad \text{and}
\]

\[
    \text{km}^{OPT}_{(X,K)} := \min_{M \subset \mathbb{R}^D \atop |M| = K} \text{km}_X(M) .
\]

Definition 13 (Ball). For \( x \in \mathbb{R}^D \), and \( r \geq 0 \), we let

\[
    B(x, r) := \{y \in \mathbb{R}^D \mid \|y - x\|_2 \leq r\} .
\]

The following lemma is well known (e.g. used in the proof of Theorem 2 in [Inaba et al., 1994]).

Lemma 14. Let \( C = \{(x_n, w_n)\}_{n \in [N]} \subset \mathbb{R}^D \times \mathbb{R}_+ \) be a weighted point set and \( \mu \in \mathbb{R}^D \). Then,

\[
    \sum_{(x_n, w_n) \in C} w_n \|x_n - \mu\|_2^2 = \text{km}(C) + w(C) \|\mu - \mu(C)\|_2^2 .
\]

Lemma 15. For all \( a, b, c \in \mathbb{R}^D \) we have

\[
    \|c - a\|_2^2 - \|c - b\|_2^2 \leq \|a - b\|_2^2 + 2 \|a - b\|_2 \|c - b\|_2
\]

Lemma 16. Let \( \epsilon \in [0, 1] \), \( c > 1 \), and \( m \in \mathbb{N} \). Then, for all \( i \in [m] \) it holds

\[
    \left(1 + \frac{\epsilon}{2mc}\right)^i \leq 1 + i \cdot \frac{\epsilon}{mc} .
\]

Lemma 17. For all \( a, b \in \mathbb{R} \) we have

1. \( 2ab \leq a^2 + b^2 \),

2. \( (a + b)^2 \leq 2(a^2 + b^2) \) and

3. \( (a + b + c)^2 \leq 3(a^2 + b^2 + c^2) \).
B Proof of Theorem $[5]$  

To prove Theorem $[5]$ we consider the following algorithm.

\textbf{Algorithm 18 (Γ-Weak Coreset).} We are given $X = \{x_n\}_{n \in [N]} \subset \mathbb{R}^D$, $K \in \mathbb{N}$, $\epsilon \in [0, 1]$, $\delta \in [0, 1]$, and $\gamma \in \mathbb{N}$.

1. Apply the randomized algorithm from [Aggarwal et al., 2009] to obtain, with probability at least $1 - \delta$, an $(\alpha, \beta)$-bicriteria approximation $A = \{a_1, \ldots, a_\eta\} \subseteq \mathbb{R}^D$ for the $K$-means problem where $\alpha, \beta = \mathcal{O}(1)$, i.e.

$$\eta := \lceil \beta K \rceil \quad \text{and} \quad \text{km}_X(A) \leq \alpha \text{km}^{OPT}_{X,K}, \quad (3)$$

2. Let

$$\nu := \lceil \log(\alpha N) \rceil \quad \text{and} \quad R := \sqrt{\frac{1}{\alpha N}} \text{km}_X(A).$$

3. For each $k \in [\eta]$, let

$$X_k := \left\{ x \in X \mid a_k = \arg \min_{a \in A} \{ \| x - a \|_2 \} \right\}.$$ 

and, for each $k \in [\eta]$ and $j \in \{0, 1, \ldots, \nu\}$, let

$$X_{k,j} := \begin{cases} X_k \cap B(a_k, R) & \text{if } j = 0 \\ X_k \cap (B(a_k, 2^j R) \setminus B(a_k, 2^{j-1} R)) & \text{if } j \geq 1 \end{cases}.$$ 

4. Let

$$M := 61 \left( \alpha K^{m-1} \right)^2 \epsilon^{-2} \ln \left( 4 \eta \nu \gamma K \delta^{-1} \right).$$

5. For each $k \in [\eta]$ and $j \in [\nu]$, uniformly sample a multiset $S_{k,j}$ from $X_{k,j}$ with $|S_{k,j}| = M$.

6. Set the weight of each point $s \in S_{k,j}$ to

$$\omega_s := |X_{k,j}|/M.$$ 

7. Let

$$S := \bigcup_{k,j} \{ (s, \omega_s) \mid s \in S_{k,j} \}.$$
Lemma 19. Let \( \epsilon, \delta > 0 \) and \( f : X \to \mathbb{R}, \ F \in \mathbb{R} \) be such that \( \forall x \in X : 0 \leq f(x) \leq F \). Let \( S \subset X \) be a uniform sample multiset of size \( |S| \geq \frac{1}{2} \epsilon^{-2} \ln (2/\delta) \). Then

\[
\Pr \left[ \left| \frac{1}{|X|} \sum_{x \in X} f(x) - \frac{1}{|S|} \sum_{s \in S} f(s) \right| \leq \epsilon F \right] \geq 1 - \delta.
\]

Proof of Theorem 5. First, we show that the property holds for some fixed \( C \subset \Gamma, |C| \leq K \), with high probability. To this end, consider arbitrary but fixed \( \{\mu_1, \ldots, \mu_L\} = C \subset \Gamma \) with \( L \leq K \).

From Lemma 3, we know that for all \( x \in X \) we have \( 0 \leq \phi^{(m)}_{\{x\}}(C) \leq km_{\{x\}} C = d(x, C)^2 \). Fix some \( k \in [\eta], j = \{0, 1, \ldots, \nu\} \) and let

\[ x^* := \arg \min_{x \in X_{k,j}} \{d(x, C)^2\} \]

Then, for each \( x_n \in X_{k,j} \) we can bound

\[
d(x_n, C)^2 \leq 2 \left( d(x^*, C)^2 + d(x^*, x_n)^2 \right) \\
\leq 4 \left( d(x^*, C)^2 + d(x^*, a_k)^2 + d(a_k, x)^2 \right) \\
\leq 4 \left( d(x^*, C)^2 + 2^{2j+1} R^2 \right).
\]

Next, we rewrite

\[
\left| \phi^{(m)}_{X_{k,j}}(C) - \phi^{(m)}_{S_{k,j}}(C) \right| = |X_{k,j}| \left| \frac{1}{X_{k,j}} \sum_{x \in X_{k,j}} \phi^{(m)}_{\{x\}}(C) - \frac{1}{S_{k,j}} \sum_{s \in S_{k,j}} \omega_s \phi^{(m)}_{\{s\}}(C) \right|
\]

\[
= |X_{k,j}| \left| \frac{1}{X_{k,j}} \sum_{x \in X_{k,j}} \phi^{(m)}_{\{x\}}(C) - \frac{1}{S_{k,j}} \sum_{s \in S_{k,j}} \phi^{(m)}_{\{s\}}(C) \right|.
\]

Then, apply Lemma 19. Since \( M \geq 1/2(11\alpha K^{-m-1})^2 \epsilon^{-2} \ln \left( 4\eta \nu |\Gamma|^K \delta^{-1} \right) \), we obtain that, with probability at least \( 1 - \delta/2(\nu K)^{-1} \), we have

\[
\left| \phi^{(m)}_{X_{k,j}}(C) - \phi^{(m)}_{S_{k,j}}(C) \right| \leq \frac{\epsilon}{11\alpha K^{-m-1}} |X_{k,j}| 4(d(x^*, C)^2 + 2^{2j+1} R^2) \quad (4)
\]

Observe that \( \{X_{k,j}\}_{k,j} \) is a partition of \( X \).

The following concentration bound was presented by [Haussler, 1992] and is crucial to our analysis of Algorithm 18.
By definition of $x^*$, we can bound 

$$|X_{k,j}| d(x^*, C)^2 \leq \sum_{x \in X_{k,j}} d(x, C)^2 = \text{km}_{X_{k,j}}(C).$$

For $j = 0$, we trivially obtain $|X_{k,j}| 2^{2j+1} R^2 = |X_{k,j}| 2R^2$. For $j \geq 1$ we have, by construction of $X_{k,j}$, that $2^{2j-2} R^2 \leq \|x - a_k\|^2 = d(x, A)^2$ for all $x \in X_{k,j}$ and hence

$$|X_{k,j}| 2^{2j+1} R^2 \leq 8 \sum_{x \in X_{k,j}} \|x - a_k\|^2 = 8 \text{km}_{X_{k,j}}(A).$$

Using these upper bounds, we can conclude from Equation (4) that

$$\left| \phi_{X_{k,j}}^{(m)}(C) - \phi_{S_{k,j}}^{(m)}(C) \right| \leq \frac{\epsilon}{11\alpha K^{m-1}} \left( \text{km}_{X_{k,j}}(C) + 8 \text{km}_{X_{k,j}}(A) + |X_{k,j}| 2R^2 \right),$$

with probability at least $1 - \delta / 2(\eta \nu |\Gamma|^K)^{-1}$.

Using the union bound, we know that this holds simultaneously for every $k \in [\eta]$ and $j = \{0, 1, \ldots, \nu\}$ with probability at least $1 - \delta / 2 |\Gamma|^{-K}$. Hence, by summing over all $k \in [\eta]$ and $j = \{0, 1, \ldots, \nu\}$, we can conclude that

$$\left| \frac{1}{|X|} \sum_{x \in X} \phi_{\{x\}}^{(m)}(C) - \frac{1}{|S|} \sum_{s \in S} \phi_{\{s\}}^{(m)}(C) \right| \leq \sum_{k=1}^{\eta} \sum_{j=0}^{\nu} \left| \phi_{X_{k,j}}^{(m)}(C) - \phi_{S_{k,j}}^{(m)}(C) \right|$$

$$\leq \frac{\epsilon}{11\alpha K^{m-1}} \left( \sum_{k=1}^{\eta} \sum_{j=0}^{\nu} \left( \text{km}_{X_{k,j}}(C) + 8 \text{km}_{X_{k,j}}(A) + |X_{k,j}| 2R^2 \right) \right)$$

$$= \frac{\epsilon}{11\alpha K^{m-1}} \left( \text{km}_X(C) + 8 \text{km}_X(A) + 2NR^2 \right)$$

$$\leq \frac{\epsilon}{11\alpha K^{m-1}} \left( \text{km}_X(C) + 8\alpha \text{km}^{OPT}_{(X,K)} + 2\text{km}^{OPT}_{(X,K)} \right)$$

$$\leq \frac{\epsilon}{K^{m-1}} \text{km}_X(C) \leq \epsilon \phi_X^{(m)}(C)$$

holds with probability at least $1 - \delta / 2 |\Gamma|^{-K}$. Analogously, we obtain that

$$\left| \frac{1}{|X|} \sum_{x \in X} \text{km}_{\{x\}}(C) - \frac{1}{|S|} \sum_{s \in S} \text{km}_{\{s\}}(C) \right| \leq \frac{\epsilon}{K^{m-1}} \text{km}_X(C),$$

also with probability at least $1 - \delta / 2 |\Gamma|^{-K}$.
Thus, we have

\[
\Pr \left[ \left| \phi^{(m)}_X(C) - \phi^{(m)}_S(C) \right| \leq \epsilon \phi^{(m)}_X(C) \right] \geq 1 - \delta/2 |\Gamma|^{-K} \tag{5}
\]

and

\[
\Pr \left[ \left| \text{km}_X(C) - \text{km}_S(C) \right| \leq \frac{\epsilon}{K^{m-1}} \text{km}_X(C) \right] \geq 1 - \delta/2 |\Gamma|^{-K} \tag{6}
\]

Recall that Equations (5) and (6) hold for arbitrary fixed \( C \subset \Gamma, |C| \leq K \). Observe that there are at most \(|\Gamma|^K\) many subsets of size not larger than \( K \) of \( \Gamma \). Hence, using union bound, we obtain that the constructed set \( S \) has the desired property, with probability at least \( 1 - \delta/2 \).

It remains to analyze the size of \( S \) and the runtime. By construction, the size of \( S \) is bounded by

\[
|S| \leq \eta \nu 61(\alpha K^{m-1})^2 \epsilon^{-2} \ln \left( 4\eta \nu|\Gamma|^K \delta^{-1} \right)
\]

\[
= \mathcal{O} \left( K^{2m-1}\epsilon^{-2} \log(N) \log(K \log(N)^{\gamma \delta^{-1}}) \right) \quad \text{(Def. of } \eta \text{ and } \nu; \ \alpha, \beta \in \mathcal{O}(1))
\]

\[
= \mathcal{O} \left( \log(N) \log \log(N) \cdot K^{2m+1}\epsilon^{-2} \log(\gamma) \log(\delta^{-1}) \right)
\]

The algorithm from [Aggarwal et al., 2009] which is applied in Step \( n \) has runtime \( \mathcal{O}(NKD \log(\delta^{-1})) \). Furthermore, for each \( x \in X \) and \( k \in [\eta] \), there is exactly one \( j \in \{0, 1, \ldots, \nu \} \) such that \( x \in X_{k,j} \). This index \( j \) can be determined in time \( \mathcal{O}(D) \). Moreover, sampling the multisets \( \{S_{k,j}\}_{k \in [\eta], j \in \{0, 1, \ldots, \nu \}} \) takes time \( \mathcal{O}(|S|) \). Hence, the overall runtime is given by \( \mathcal{O}(NKD \log(\delta^{-1}) + |S|) \). This yields the theorem. \( \square \)

## C Proof of Theorem 7

### C.1 Preliminaries

To prove Theorem 7, we consider Algorithm 18 which we apply to \( X, K, \tilde{\epsilon} := \frac{\alpha}{160K^{m-1}} \), \( \delta := \delta/2 \), and

\[
\gamma := \left( \log(\alpha N) + m \log \left( \frac{64\alpha m K^2}{\tilde{\epsilon}} \right) \right) \cdot K \cdot \left( \frac{12bK}{\tilde{\epsilon}} \right)^D,
\]

where \( \alpha \) is given by he \( (\alpha, \beta) \)-bicriteria approximation algorithm used in the first step of Algorithm 18.
Additionally, we consider the following quantities:

\[
\Phi := \left\lceil \frac{1}{2} \left( \log (\alpha N) + m \cdot \log \left( \frac{64amK^2}{\epsilon} \right) \right) \right\rceil, \tag{7}
\]

\[
R := \sqrt{\frac{\max_X(A)}{\alpha N}}, \tag{8}
\]

\[
U := \bigcup_{k=1}^{K} B(a_k, 2^{\Phi} R), \tag{9}
\]

\[
\kappa := \alpha K^{m-1}, \quad \text{and}
\]

\[
b := 864. \tag{10}
\]

For all \( k \in [K] \) and \( j \in \{0, \ldots, \Phi\} \), we set

\[
L_{k,j} := \begin{cases} B(a_k, R), & \text{if } j = 0 \\ B(a_k, 2^{j} R) \setminus B(a_k, 2^{j-1} R) & \text{if } j \geq 1 \end{cases}, \tag{12}
\]

For each \( k \in [K] \) and \( j \in \{0, \ldots, \Phi\} \), construct an axis-parallel grid with side length

\[
\rho_j = \frac{2^j \epsilon R}{\kappa \sqrt{D}}
\]

to partition \( L_{k,j} \) into cells. Pick the center of each cell as a representative of the points inside the cell. Then, we call

\[
G := \{x' \in \mathbb{R}^D \mid \exists x \in U: x' \text{ is the representative of } x\}
\]

be the set of all representatives.

### C.2 Results from [Blömer et al., 2015](#)

**Theorem 20.** Let \( C \subseteq U \). Let \( C' \) be the representatives of points in \( C \), i.e. \( C' := \{\mu' \in G \mid \exists \mu \in C: \mu' \text{ is the representative of } \mu\} \). Then,

\[
\left| \phi_X^{(m)}(C) - \phi_X^{(m)}(C') \right| \leq \frac{\epsilon}{4} \phi_X^{(m)}(C).
\]

**Lemma 21.** For each \( \mu \in U \) with representative \( \mu' \in G \) it holds

\[
\|\mu - \mu'\|_2 \leq \frac{2\epsilon}{b\kappa} (d(\mu, A) + R) \leq \frac{2\epsilon}{b\kappa} (\|x - \tilde{\mu}\|_2 + d(x, A) + R).
\]

and

\[
\|\mu - \mu'\|_2^2 \leq \frac{12\epsilon^2}{b^2\kappa^2} \left( \|x - \tilde{\mu}\|_2^2 + d(x, A)^2 + R^2 \right)
\]

for all \( x \in \mathbb{R}^D \) and \( \tilde{\mu} \in \{\mu, \mu'\} \).
Theorem 22. Let \( C = \{\mu_k\}_{k \in [K]} \subset U \). Let \( C' \) be the representatives of points in \( C \), i.e. \( C' := \{\mu' \in G \mid \exists \mu \in C : \mu' \text{ is the representative of } \mu\} \).

For all \( x_n \in X \), we have
\[
d(x_n, C')^2 \leq d(x_n, C)^2 + \frac{18\epsilon}{bK} (2d(x_n, C)^2 + d(x_n, A)^2 + R^2) .
\]

Theorem 23. Let \( C \subseteq U \). Let \( C' \) be the representatives of points in \( C \), i.e. \( C' := \{\mu' \in G \mid \exists \mu \in C : \mu' \text{ is the representative of } \mu\} \). Then, for all \( x_n \in X \), we have
\[
\left| \phi_{\{x_n\}}^{(m)}(C) - \phi_{\{x_n\}}^{(m)}(C') \right| \leq \frac{72\epsilon}{bK} (d(x_n, C)^2 + d(x_n, A)^2 + R^2) .
\]

Observation 24. We have
\[
N \cdot R^2 \leq \sum_{n=1}^{N} d(x_n, A)^2 = \text{km}_X(A) \quad \text{ (Definition 12, Equation (8))}
\leq \alpha \cdot \text{km}^{\text{OPT}}_{(X,K)} \quad \text{ (Equation (3))}
\leq \alpha K^{m-1} \phi^{\text{OPT}}_{(X,K,m)} \quad \text{ (Lemma 3)}
= \kappa \cdot \phi^{\text{OPT}}_{(X,K,m)},
\]

C.3 Analysis

Analogously to [Chen, 2009] one can show that the following holds true

Observation 25.
\[
|G| = (\log(|X|))^K O(mD \log(1/\epsilon)) .
\]

Hence, \(|G| \leq \gamma\). Thus, by Theorem 5 with probability at least \(1 - \delta/2\), \( S \) is a \( G \)-weak \((K, \hat{\epsilon})\)-coreset. In the following, we assume that \( S \) is indeed a \( G \)-weak \((K, \hat{\epsilon})\)-coreset, i.e. for all \( C' \subseteq G \), \(|C'| = K\), we have
\[
\left| \phi_X^{(m)}(C') - \phi_S^{(m)}(C') \right| \leq \hat{\epsilon} \phi_X^{(m)}(C') \quad . \tag{14}
\]

In the following, we fix an arbitrary solution \( C = \{\mu_k\}_{k \in [L]} \subseteq \mathbb{R}^D \) with respect to \( X \) with \( L \leq K \) and
\[
\min_l \ R_l \geq \left( \frac{\epsilon}{4mK^2} \right)^m, \tag{15}
\]
where
\[
R_l = \sum_{n=1}^{N} r_{nl}^m .
\]

The following claims and their proofs are an analogon of claims in [Chen, 2009].
Claim 26. If there exists some $l \in [L]$ such that $\mu_l \in C$ and $\mu_l \notin U$, then

$$\left| \phi^{(m)}_X(C) - \phi^{(m)}_S(C) \right| \leq \epsilon \phi^{(m)}_X(C).$$

Proof. For each $x \in X$ let $a(x) = \arg \min_{a \in A} \|a - x\|_2$. Then, $d(A, x)^2 = \|a(x) - x\|_2^2 \leq \kappa m X(A) = N \cdot R^2$. Furthermore, since $\mu_l \notin U$, we know that $\|\mu_l - a\|_2 \geq 2^\Phi R$. By the triangle inequality, we conclude

$$\|\mu_l - x\|_2 \geq \|\mu_l - a\|_2 - \|a - x\|_2 \geq 2^\Phi R - \sqrt{N} \cdot R = (2^\Phi - \sqrt{N})R.$$

Using Lemma 3, we obtain

$$\phi^{(m)}_X(C) \geq \sum_{n=1}^{N} r_{nl}^m \|x - \mu_l\|_2^2 \geq \left( \sum_{n=1}^{N} r_{nl}^m \right) (2^\Phi - \sqrt{N})^2 R^2 \geq R_l (2^\Phi - \sqrt{N})^2 \frac{\kappa m X(A)}{\alpha N}.$$

Hence,

$$\kappa m X(A) \leq \frac{\alpha N}{R_l (2^\Phi - \sqrt{N})^2} \cdot \phi^{(m)}_X(C). \quad (16)$$

In the following, we denote representative of $x_n \in X$ in $S$ by $\tilde{x}_n$. Let $\{r_{nl}\}_{n,l}$ be the optimal responsibilities induced by $C$ with respect to the $\{\tilde{x}_n\}_n$. Recall that, by construction, the weight $\omega_{\tilde{x}}$ of $\tilde{x}$ equals the number
of points in $X$ whose representative is $\tilde{x}$ (cf. Algorithm 18). Thus, we have

$$\left| \phi_{X}^{(m)}(C) - \phi_{S}^{(m)}(C) \right|$$

$$\leq \max \left\{ \sum_{l=1}^{L} \sum_{n=1}^{N} (\tilde{r}_{nl})^{m} \left( \|x_{n} - \mu_{l}\|^{2} - \|\tilde{x}_{n} - \mu_{l}\|^{2} \right), \sum_{l=1}^{L} \sum_{n=1}^{N} r_{nl}^{m} \left( \|\tilde{x}_{n} - \mu_{l}\|^{2} - \|x_{n} - \mu_{l}\|^{2} \right) \right\}$$

(by optimality of responsibilities)

$$\leq \sum_{n=1}^{N} \|x_{n} - \tilde{x}_{n}\|^{2}$$

$$+ 2 \max \left\{ \sum_{n=1}^{N} \sum_{l=1}^{L} (\tilde{r}_{nl})^{m} \|\tilde{x}_{n} - \mu_{l}\| \|x_{n} - \tilde{x}_{n}\|, \sum_{l=1}^{L} \sum_{n=1}^{N} r_{nl}^{m} \|x_{n} - \mu_{l}\| \|x_{n} - \tilde{x}_{n}\| \right\},$$

where in the last inequality we use Lemma 15 and the fact that $\sum_{n=1}^{N} r_{nl}^{m} \leq 1$ and $\sum_{n=1}^{N} (\tilde{r}_{nl})^{m} \leq 1$ for each $l \in [L]$.

By the binomial formulas, for all $a, x, y \in \mathbb{R}, a \neq 0$, we have $2axy \leq a^{2}x^{2} + \frac{1}{a^{2}}y^{2}$. Hence, for each $\hat{\epsilon} \in [0, 1]$,

$$\sum_{l=1}^{L} \sum_{n=1}^{N} (\tilde{r}_{nl})^{m} \|\tilde{x}_{n} - \mu_{l}\| \|x_{n} - \tilde{x}_{n}\|$$

$$\leq \sum_{l=1}^{L} \sum_{n=1}^{N} (\tilde{r}_{nl})^{m} \left( \frac{\hat{\epsilon}}{2} \|\tilde{x}_{n} - \mu_{l}\|^{2} + \frac{2}{\hat{\epsilon}} \|x_{n} - \tilde{x}_{n}\|^{2} \right)$$

$$\leq \frac{\hat{\epsilon}}{2} \phi_{S}^{(m)}(C) + \frac{2}{\hat{\epsilon}} \sum_{n=1}^{N} \|x_{n} - \tilde{x}_{n}\|^{2}.$$  \hspace{1cm} (since $\sum_{n=1}^{N} (\tilde{r}_{nl})^{m} \leq 1$ for each $l \in [L]$)

Thus,

$$\sum_{l=1}^{L} \sum_{n=1}^{N} r_{nl}^{m} \|x_{n} - \mu_{l}\| \|x_{n} - \tilde{x}_{n}\|$$

$$\leq \frac{\hat{\epsilon}}{2} \phi_{X}^{(m)}(C) + \frac{2}{\hat{\epsilon}} \sum_{n=1}^{N} \|x_{n} - \tilde{x}_{n}\|^{2}.$$  \hspace{1cm} (since $\sum_{n=1}^{N} (\tilde{r}_{nl})^{m} \leq 1$ for each $l \in [L]$)

Analogously,

$$\left| \phi_{X}^{(m)}(C) - \phi_{S}^{(m)}(C) \right| \leq \left( 1 + \frac{4}{\hat{\epsilon}} \right) \sum_{n=1}^{N} \|x_{n} - \tilde{x}_{n}\|^{2} + \hat{\epsilon} \cdot \max \left\{ \phi_{X}^{(m)}(C), \phi_{S}^{(m)}(C) \right\}.$$  \hspace{1cm} (17)
From [Chen, 2009], we know that $\sum_{n=1}^{N} \| x_n - \tilde{x}_n \|^2 \leq 20 \cdot k_{X}(A)$. With Equation (16), we can conclude

$$\sum_{n=1}^{N} \| x_n - \tilde{x}_n \|^2 \leq 20 \cdot \frac{\alpha N}{R_t(2^\Phi - \sqrt{N})^2} \cdot \phi_X^{(m)}(C).$$

(18)

Furthermore, we know from Lemma 3, that $\phi_S^{(m)}(C) \leq k_{S}(C)$. From Theorem 5, we know that $k_{S}(C) \leq (1 + \tilde{\epsilon}) k_{X}(C)$. Hence, due to Lemma 3, we can conclude

$$\phi_S^{(m)}(C) \leq (1 + \tilde{\epsilon}) L^{m-1} \phi_X^{(m)}(C) \leq (1 + \tilde{\epsilon}) K^{m-1} \phi_X^{(m)}(C).$$

(19)

By plugging Equation (18) and (19) into Equation (17), we obtain that

$$\left| \phi_X^{(m)}(C) - \phi_S^{(m)}(C) \right| \leq \left( 1 + \frac{4}{\tilde{\epsilon}} \right) \frac{20 \cdot \alpha N}{R_t(2^\Phi - \sqrt{N})^2} \phi_X^{(m)}(C) + \epsilon (1 + \tilde{\epsilon}) K^{m-1} \phi_X^{(m)}(C).$$

Observe that

$$\left( 2^\Phi - \sqrt{N} \right)^2 = \left( \sqrt{\alpha N \cdot \left( \frac{64 \alpha m K^2}{\tilde{\epsilon}} \right)^m} - \sqrt{N} \right)^2$$

$$= N \left( \sqrt{\alpha \cdot \left( \frac{64 \alpha m K^2}{\tilde{\epsilon}} \right)^m} - 1 \right)^2$$

$$= N \left( \alpha \cdot \left( \frac{64 \alpha m K^2}{\tilde{\epsilon}} \right)^m - 2 \sqrt{\alpha \cdot \left( \frac{64 \alpha m K^2}{\tilde{\epsilon}} \right)^m} \right) + 1$$

$$\geq N \left( \alpha \cdot \left( \frac{1}{2} \cdot \frac{64 \alpha m K^2}{\tilde{\epsilon}} \right)^m \right) + 1$$

$$\geq N \alpha \cdot \left( \frac{32 \alpha m K^2}{\tilde{\epsilon}} \right)^m,$$

(20)

(21)

where the second to last inequality is due to the fact that, for all $y \in \mathbb{R}$ with $y \geq 16$, we have $\sqrt{y} \leq y/4$ and $\alpha \cdot \left( \frac{64 \alpha m K^2}{\tilde{\epsilon}} \right)^m \geq 128^2$ since $\alpha \geq 1$, $m \geq 2$, $K \geq 1$, $\epsilon \leq 1.$
Hence, with \( \hat{\epsilon} := \tilde{\epsilon} = \frac{\epsilon}{80K^{m-1}} \), we obtain

\[
\left(1 + \frac{4}{\hat{\epsilon}}\right) 20 \frac{\alpha N}{R_k(2^\Phi - \sqrt{N})^2} + \hat{\epsilon}(1 + \hat{\epsilon})K^{m-1}
\]

\[
\leq \left(1 + \frac{4}{\hat{\epsilon}}\right) 20 \frac{\alpha N(4mK^2)^m}{\epsilon^m(2^\Phi - \sqrt{N})^2} + \hat{\epsilon}(1 + \hat{\epsilon})K^{m-1}
\]  

(Equation \(15\))

\[
\leq \left(1 + \frac{4}{\hat{\epsilon}}\right) 20 \frac{(4mK^2)^m}{\epsilon^m \cdot \left(\frac{32\alpha mK^2}{\hat{\epsilon}}\right)^m} + \hat{\epsilon}(1 + \hat{\epsilon})K^{m-1}
\]  

(Equation \(20\))

\[
\leq \left(1 + \frac{4}{\hat{\epsilon}}\right) 20 \left(\frac{\epsilon}{\hat{\epsilon}}\right)^m + \hat{\epsilon}(1 + \hat{\epsilon})K^{m-1}
\]  

\[
= \left(1 + \frac{640K^{m-1}}{\epsilon^3}\right) 20 \left(\frac{\epsilon^2}{160K^{m-1}}\right)^m + \frac{\epsilon^3}{160} \left(1 + \frac{\epsilon^3}{160K^{m-1}}\right)
\]  

(Def. \(\tilde{\epsilon}\))

\[
\leq \frac{20\epsilon^2m}{(160K^{m-1})^m} + \frac{20 \cdot 640K^{m-1} \epsilon^{2m}}{(160K^{m-1})^m \epsilon^3} + \frac{\epsilon}{2}
\]

\[
\leq \frac{\epsilon}{4} + \frac{\epsilon}{2K^{m-1}} + \frac{\epsilon}{2}
\]  

\[
\leq \epsilon.
\]

\[\square\]

Claim 27.

\[
km_S(A) \leq 5\kappa\phi_X^{(m)}(C)
\]

Proof. By definition of Algorithm 18 we have

\[
km_S(A) = \sum_{(s, \omega_s) \in S} \omega_s \min_{a \in A} \| s - a \|^2
\]

\[
= \sum_{k=1}^{K} \sum_{j=1}^{\nu} \sum_{(s, \omega_s) \in S_{k,j}} \omega_s \min_{a \in A} \| s - a \|^2
\]

\[
\leq \sum_{k=1}^{K} \sum_{j=1}^{\nu} \sum_{(s, \omega_s) \in S_{k,j}} \omega_s \| s - a_k \|^2
\]

\[
\leq \sum_{k=1}^{K} \sum_{j=1}^{\nu} \left( \sum_{s \in X: s \in S_{k,j}} \omega_s \right) (2^j R)^2
\]

\( (s \in S_{k,j}) \)
Recall the definition of $\omega_s$ from Algorithm 18. We can conclude that

$$\text{km}_S(A) \leq \sum_{k=1}^{K} \sum_{j=1}^{\nu} |S_{k,j}| \cdot \frac{|X_{i,j}|}{M} (2^j R)^2$$

$$\leq \sum_{k=1}^{K} \sum_{j=1}^{\nu} |X_{k,j}| (2^j R)^2$$

$$\leq \sum_{k=1}^{K} \sum_{j=1}^{\nu} \sum_{x_n \in X_{k,j}} (2^j R)^2$$

$$\leq \sum_{k=1}^{K} \sum_{j=1}^{\nu} \sum_{x_n \in X_{k,j}} (4d(A, x_n)^2 + R^2)$$

(analogously to Claim 21)

$$\leq \sum_{x_n \in X} (4d(A, x_n)^2 + R^2)$$

$$\leq 4 \text{km}_X(A) + N R^2$$

$$\leq 5\kappa \cdot \phi^{OPT}_{(X,K,m)}$$

(Observation 24)

$$\leq 5\kappa \cdot \phi^{(m)}_X(C) .$$

(|$C| = L \leq K$)

Claim 28. If $C \subseteq \mathcal{U}$,

$$\text{km}_S(C) - \text{km}_X(C) \leq 4\tilde{\epsilon} \phi_X(C)$$

Proof. Let $C' \subseteq \mathcal{G}$ be the representatives of $C \subseteq \mathcal{U}$. Observe that

$$\text{km}_S(C) - \text{km}_X(C)$$

$$\leq (\text{km}_S(C) - \text{km}_S(C')) + |\text{km}_S(C') - \text{km}_X(C')| + (\text{km}_X(C') - \text{km}_X(C)) .$$

From Lemma 22 we know that for all $x_n \in X \subset \mathbb{R}^D$,

$$d(x_n, C')^2 - d(x_n, C)^2 \leq \frac{18\tilde{\epsilon}}{b\kappa} \left((2d(x_n, C')^2 + d(x_n, A)^2 + R^2) .$$
By summing over all \( x_n \in X \subseteq \mathbb{R}^D \), we obtain

\[
\begin{align*}
\kappa m_X(C') - \kappa m_X(C) & \leq \frac{18\tilde{\varepsilon}}{b\kappa} (2\kappa m_X(C) + \kappa m_X(A) + NR^2) \\
& \leq \frac{18\tilde{\varepsilon}}{b\kappa} \cdot 2(1 + \alpha) \kappa m_X(C) \quad \text{(Equation 3 and 8)} \\
& \leq \frac{18\tilde{\varepsilon}}{bK^{m-1}} \cdot 4 \cdot \kappa m_X(C) \quad \text{(Equation 10)} \\
& \leq \frac{\tilde{\varepsilon}}{K^{m-1}} \kappa m_X(C) \quad \text{(Equation 11, \( b \geq 72 \))}
\end{align*}
\]

Hence,

\[
\kappa m_X(C') - \kappa m_X(C) \leq \frac{\tilde{\varepsilon}}{K^{m-1}} \kappa m_X(C) . \tag{22}
\]

Using Theorem 5, we can conclude

\[
\begin{align*}
|\kappa m_X(C') - \kappa m_S(C')| & \leq \frac{\tilde{\varepsilon}}{K^{m-1}} \kappa m_X(C') \\
& \leq \frac{\tilde{\varepsilon}}{K^{m-1}} (1 + \frac{\tilde{\varepsilon}}{K^{m-1}}) \kappa m_X(C) \\
& \leq \frac{2\tilde{\varepsilon}}{K^{m-1}} \kappa m_X(C) .
\end{align*}
\]

From Lemma 22, we know that for all \( x_n \in X \subseteq \mathbb{R}^D \), we have

\[
d(x_n, C')^2 - d(x_n, C)^2 \leq \frac{18\varepsilon}{b\kappa} (2d(x_n, C)^2 + d(x_n, A)^2 + R^2) .
\]

By summing over all \( (s, \omega_s) \in S \subseteq X \times \mathbb{R} \), we obtain

\[
\begin{align*}
\kappa m_S(C) - \kappa m_S(C') & \leq \sum_{(s, \omega_s) \in S} \kappa m_{\{s\}}(C') - \kappa m_{\{s\}}(C) \\
& \leq \frac{18\tilde{\varepsilon}}{b\kappa} \sum_{(s, \omega_s) \in S} (2d(x_n, C)^2 + d(x_n, A)^2 + R^2) \\
& \leq \frac{18\tilde{\varepsilon}}{b\kappa} (2\kappa m_S(C) + \kappa m_S(A) + NR^2) \quad \text{(Def. \( \omega_s \))} \\
& \leq \frac{18\tilde{\varepsilon}}{b\kappa} (2\kappa m_S(C) + \kappa m_S(A) + \alpha \kappa m_X(C)) \quad \text{(Equation 8 and 3)} \\
& \leq \frac{18\tilde{\varepsilon}}{b\kappa} (2\kappa m_S(C) + 6\alpha \kappa m_X(C)) . \quad \text{(Claim 27)}
\end{align*}
\]
From Theorem 5, we know that
$$|km_X(C) - km_S(C)| \leq \frac{\epsilon}{K^m - 1} km_X(C).$$

Hence,
$$km_S(C) - km_S(C') \leq \frac{18\hat{\epsilon}}{b\kappa} (2 km_S(C) + 6\alpha km_X(C))$$
$$\leq \frac{18\hat{\epsilon}}{b\kappa} \left( \frac{2}{K^m - 1} km_X(C) + 6\alpha km_X(C) \right)$$
$$\leq \frac{126\hat{\epsilon}\alpha}{b\kappa} km_X(C)$$
$$\leq \frac{2\hat{\epsilon}}{K^m - 1} km_X(C).$$ (Equation (10) and (11))

Combining all these inequalities yields
$$km_S(C) - km_X(C) \leq (km_S(C) - km_S(C')) + |km_S(C') - km_X(C')| + (km_X(C') - km_X(C))$$
$$\leq \left( \frac{\hat{\epsilon}}{K^m - 1} + \frac{2\hat{\epsilon}}{K^m - 1} + \frac{\hat{\epsilon}}{K^m - 1} \right) km_X(C)$$
$$\leq \frac{4\hat{\epsilon}}{K^m - 1} km_X(C)$$
$$\leq 4\hat{\epsilon} \phi_X(C),$$ (Lemma 3)

which yields the claim.

\[\Box\]

**Claim 29.** If \(C \subseteq U\), then
$$\left| \phi_X^{(m)}(C) - \phi_S^{(m)}(C) \right| \leq \epsilon \phi_X^{(m)}(C).$$

**Proof.** Let \(C' \subseteq G\) be the representatives of \(C \subseteq U\). Observe that
$$\left| \phi_X^{(m)}(C) - \phi_S^{(m)}(C) \right|$$
$$\leq \left| \phi_X^{(m)}(C) - \phi_X^{(m)}(C') \right| + \left| \phi_X^{(m)}(C') - \phi_S^{(m)}(C') \right| + \left| \phi_S^{(m)}(C') - \phi_S^{(m)}(C) \right|.$$

From Theorem 20, we can conclude that
$$\left| \phi_X^{(m)}(C) - \phi_X^{(m)}(C') \right| \leq \frac{\hat{\epsilon}}{4} \phi_X^{(m)}(C).$$ (23)
Using Equation (14), we can conclude

\[
|\phi_{X}^{(m)}(C') - \phi_{S}^{(m)}(C')| \leq \tilde{c} \phi_{X}^{(m)}(C) \leq \tilde{c} \left(1 + \frac{\tilde{c}}{4}\right) \phi_{X}^{(m)}(C) \leq \frac{5}{4} \tilde{c} \cdot \phi_{X}^{(m)}(C). \tag{24}
\]

Recall that \( S \subset X \times \mathbb{R} \). From Lemma 23, we know that for all \( x_n \in X \subset \mathbb{R}^D \), we have

\[
|\phi_{\{x_n\}}^{(m)}(C) - \phi_{\{x_n\}}^{(m)}(C')| \leq \frac{18\epsilon}{b\kappa} \left(\phi_{\{x_n\}}^{(m)}(C) + 2d(x_n, A)^2 + 2R^2\right).
\]

By summing over all \((s, \omega_s) \in S \subset X \times \mathbb{R}\), we obtain

\[
\left|\phi_{S}^{(m)}(C') - \phi_{S}^{(m)}(C)\right| \leq \sum_{(s, \omega_s) \in S} \omega_s \frac{18\epsilon}{b\kappa} \left(\phi_{\{s\}}^{(m)}(C) + 2km\{s\}(A) + 2R^2\right)
\leq \frac{18\epsilon}{b\kappa} \left(\sum_{(s, \omega_s) \in S} \omega_s \phi_{\{s\}}^{(m)}(C) + 2 \sum_{(s, \omega_s) \in S} \omega_s km\{s\}(A) + 2 \left(\sum_{(s, \omega_s) \in S} \omega_s\right) R^2\right)
\leq \frac{18\epsilon}{b\kappa} \left(\phi_{S}^{(m)}(C) + 2km(S)(A) + 2NR^2\right)
\leq \frac{18\epsilon}{b\kappa} \left(\phi_{S}^{(m)}(C) + 2km(S)(A) + 2km_X(A)\right). \tag{Equation 8}
\]

Hence,

\[
\left|\phi_{S}^{(m)}(C') - \phi_{S}^{(m)}(C)\right| \leq \frac{18\epsilon}{b\kappa} \left(\phi_{S}^{(m)}(C) + 2km(S)(A) + 2km_X(A)\right). \tag{25}
\]

Recall that \( km(S)(A) \leq 5\alpha \phi_{X}^{(m)}(C) \). Furthermore, observe that, by Lemma 3 and Claim 28, \( \phi_{S}^{(m)}(C) \leq km(S) \leq (1 + 4\tilde{c}) \phi_{X}(C) \). Combining all these
inequalities yields

\[
\left| \phi_X^{(m)}(C) - \phi_S^{(m)}(C) \right| \\
\leq \left| \phi_X^{(m)}(C) - \phi_X^{(m)}(C') \right| + \left| \phi_X^{(m)}(C') - \phi_s^{(m)}(C') \right| + \left| \phi_s^{(m)}(C') - \phi_s^{(m)}(C) \right| \\
\leq \frac{\tilde{\varepsilon}}{4} \phi_X^{(m)}(C) + \frac{5}{4} \tilde{\varepsilon} \cdot \phi_X^{(m)}(C) + \frac{18\tilde{\varepsilon}}{b\kappa} \left( \phi_s^{(m)}(C) + 2km_s(A) + 2km_X(A) \right) \\
\leq \frac{\tilde{\varepsilon}}{4} \phi_X^{(m)}(C) + \frac{5}{4} \tilde{\varepsilon} \cdot \phi_X^{(m)}(C) + \frac{18\tilde{\varepsilon}}{b\kappa} \left( (1 + 4\tilde{\varepsilon})\phi_X(C) + 2km_s(A) + 2km_X(A) \right) \\
\leq \left( \frac{6}{4} \tilde{\varepsilon} + \frac{18\tilde{\varepsilon}}{b\kappa} (1 + 4\tilde{\varepsilon}) \right) \cdot \phi_X^{(m)}(C) + \frac{18\tilde{\varepsilon}}{b\kappa} \left( 2km_s(A) + 2km_X(A) \right) \\
\leq \left( \frac{6}{4} \tilde{\varepsilon} + \frac{18\tilde{\varepsilon}}{b\kappa} (1 + 4\tilde{\varepsilon}) + \frac{180\varepsilon}{b} + \frac{36\tilde{\varepsilon}}{b} \right) \cdot \phi_X^{(m)}(C) \\
\leq \left( \frac{6}{4} + \frac{36}{b\kappa} \frac{216}{b} \right) \cdot \tilde{\varepsilon} \cdot \phi_X^{(m)}(C) \\
\leq \left( \frac{6}{4} + \frac{252}{b\kappa} \right) \cdot \tilde{\varepsilon} \cdot \phi_X^{(m)}(C) \\
\leq 3 \cdot \tilde{\varepsilon} \cdot \phi_X^{(m)}(C) \\
\leq \varepsilon \cdot \phi_X^{(m)}(C) .
\]

Combining Claim 26 and 29 yields that \((1 - \varepsilon)\phi_X^{(m)}(C) \leq \phi_s^{(m)}(C) \leq (1 + \varepsilon)\phi_X^{(m)}(C)\).

It remains to analyze the size of \(S\) and the algorithms’ runtime. Observe that by definition,

\[
\log(\gamma) \subseteq O \left( \log(\log(N)) \cdot D \cdot m \log(m) \cdot \log(K) \cdot \log(\varepsilon^{-1}) \right) .
\]

From Theorem 5, we know that we can bound the runtime by

\[
O \left( ND \cdot K^{2m+1} \varepsilon^{-2} \log(\gamma) \log(\delta^{-1}) \right) \\
\subseteq O \left( N \log(\log(N)) \cdot D^2 \cdot K^{4m} \cdot m^2 \cdot \varepsilon^{-7} \cdot \log(\delta^{-1}) \right) .
\]

Analogously, using Theorem 5, we can bound the size of \(S\) by

\[
|S| \subseteq O \left( \log(N) \log(\log(N))^2 \cdot D \cdot K^{4m} \cdot m^2 \cdot \varepsilon^{-7} \cdot \log(\delta^{-1}) \right) .
\]
\section*{D Proof of Theorem 8}

In this section, we show that $(K, \epsilon)$-coresets for balanced solutions preserve approximations. That is, a good approximation for the fuzzy $K$-means problem with respect to the coreset is also a good approximation to the fuzzy $K$-means problem with respect to $X$. We already formalized this result in Theorem 8.

To prove Theorem 8, we need the following results.

\textbf{Lemma 30.} Let $X \subset \mathbb{R}^D$, $C \subset \mathbb{R}^D$, $K = |C|$, and $S \subset X \times \mathbb{R}$. Let \( \{r(s, k)\}_{s \in S, k \in [K]} \) be the optimal memberships induced by \( C \) with respect to the weighted set \( S \). Then there is a set of optimal memberships \( \{r(x, k)\}_{x \in X, k \in [K]} \) induced by \( C \) with respect to the unweighted set \( X \) where for all \( s = (x, w) \in S \) we have \( r(s, k) = r(x, k) \).

\textit{Proof.} The claim follows directly from the observation that the optimal responsibility of each point only depends on the point itself and the given set of means (cf. Equation (1)). \hfill \Box

\textbf{Corollary 31.} Let $X = \{x_n\}_{n \in [N]} \subset \mathbb{R}^D$, $m \in \mathbb{N}$, $K \in \mathbb{N}$, and $\epsilon \in [0, 1]$. Let \( C = \{\mu_k\}_{k \in [K]} \subset \mathbb{R}^D \).

There exists \( L \subseteq [K] \) such that for each \( k \in L \) we have that there exists an \( n \in [N] \) such that \( r_{nk} \geq \frac{\epsilon}{4mnK^2} \).

\textit{Proof.} can be found in [Blömer et al., 2015]. \hfill \Box

\textit{Proof of Theorem 8.} Let $\tilde{\epsilon} = \epsilon/14$. Consider an arbitrary but fixed $C \subset \mathbb{R}^D$, $|C| = K$ with

\begin{equation}
\phi_S^{(m)}(C) \leq \alpha \cdot \phi_{OPT}^{(S,K,m)} .
\end{equation}

By Corollary 31, there is an index set \( L \subseteq [K] \) such that the set \( \tilde{C} = \{\mu_k \in C | k \in L\} \) fulfills

\begin{equation}
\phi_S^{(m)}(\tilde{C}) \leq (1 + \tilde{\epsilon})\phi_S^{(m)}(C) .
\end{equation}

and such that, for every $k \in L$, there exists an $x_{n(l)} \in S$ with $\tilde{r}_{n(l)k} \geq \frac{\tilde{\epsilon}}{4mnK^2}$, where \( \{\tilde{r}_{nk}\}_{n,k} \) are the optimal memberships induced by $\tilde{C}$ with respect to $X$. \hfill \Box
S. With Lemma 30, we can conclude that for every \( k \in L \) we also have that 
\[ r_{n(k)} \geq \frac{\bar{\epsilon}}{4mK^2}, \]
where \( \{r_{nk}\}_{n,k} \) are the optimal memberships induced by \( \tilde{C} \) with respect to \( X \). Hence, for all \( k \in L \), we have 
\[ R_k = \sum_{n=1}^{N} r_{nk}^m \geq \frac{\bar{\epsilon}}{4mK^2}. \]
That means, we can apply Theorem 7 to obtain 
\[ (1 - \bar{\epsilon})\phi_X^{(m)}(\tilde{C}) \leq \phi_S^{(m)}(\tilde{C}) \leq (1 + \bar{\epsilon})\phi_X^{(m)}(\tilde{C}). \] (28)

Due to Corollary 31, there exists a set of means \( \tilde{C}_{\text{opt}}^X \subseteq C_{\text{opt}}^X \) with 
\[ \phi_X^{(m)}(\tilde{C}_{\text{opt}}^X) \leq (1 + \bar{\epsilon})\phi^{\text{OPT}}_{(X,K,m)}, \] (29)
and where for the optimal memberships \( r_{nk}^{\text{opt}} \) induced by \( \tilde{C}_{\text{opt}}^X \) it holds that for all \( k \) we have 
\[ R_k^{\text{opt}} = \sum_{n=1}^{N} (r_{nk}^{\text{opt}})^m \geq (\frac{\bar{\epsilon}}{4mK^2})^m, K. \] Hence, we can apply Theorem 7 to obtain 
\[ (1 - \bar{\epsilon})\phi_X^{(m)}(\tilde{C}_{\text{opt}}^X) \leq \phi_S^{(m)}(\tilde{C}_{\text{opt}}^X) \leq (1 + \bar{\epsilon})\phi_X^{(m)}(\tilde{C}_{\text{opt}}^X). \] (30)

Putting it all together,
\[ \phi_X^{(m)}(C) \leq \phi_X^{(m)}(\tilde{C}) \leq \frac{1}{1 - \bar{\epsilon}}\phi_S^{(m)}(\tilde{C}) \leq \frac{1 + \bar{\epsilon}}{1 - \bar{\epsilon}}\phi_S^{(m)}(C) \leq \alpha \cdot \frac{1 + \bar{\epsilon}}{1 - \bar{\epsilon}}\phi_X^{\text{OPT}}_{(X,K,m)} \leq \alpha \cdot \frac{1 + \bar{\epsilon}}{1 - \bar{\epsilon}}\phi_S^{(m)}(\tilde{C}_{\text{opt}}^X) \leq \alpha \cdot \frac{1 + \bar{\epsilon}}{1 - \bar{\epsilon}}\phi_X^{(m)}(\tilde{C}_{\text{opt}}^X) \leq \alpha \cdot \frac{1 + \bar{\epsilon}^2}{1 - \bar{\epsilon}}\phi_X^{(m)}(\tilde{C}_{\text{opt}}^X) \leq \alpha \cdot \frac{(1 + \bar{\epsilon})^2}{1 - \bar{\epsilon}}\phi_X^{(m)}(\tilde{C}_{\text{opt}}^X). \] (28)

Using Lemma 16 and the fact that \( \bar{\epsilon} = \epsilon/14 \leq 0.5 \), we can conclude
\[ \frac{(1 + \bar{\epsilon})^2}{1 - \bar{\epsilon}} \leq \frac{1 + 6\bar{\epsilon}}{1 - \bar{\epsilon}} = 1 + \frac{7\bar{\epsilon}}{1 - \bar{\epsilon}} \leq 1 + 14\bar{\epsilon} = 1 + \epsilon. \]
This yields the claim. \( \square \)
E Proof of Corollary 11

By combining the dimension reduction technique by Johnson-Lindenstrauss and our coreset construction from Theorem 7 and our result from Theorem 8, we obtain the following result.

Theorem 32. There is an algorithm that, given $X \subset \mathbb{R}^D$, $K \in \mathbb{N}$ and $\epsilon \in (0, 1/2]$, computes a set $S \subset \mathbb{R}^D$ and a mapping $\pi : \mathbb{R}^D \rightarrow \mathbb{R}^{\tilde{D}}$, where

$$|S| \in \mathcal{O}(\log(N) \log(\log(N))^2 \cdot D \cdot K^{4m} \cdot m^2 \cdot \epsilon^{-7})$$

and

$$\tilde{D} \in \mathcal{O}\left((\log(\log(N)) + \log(D) + 4m \log(K)) / \epsilon^3\right),$$

satisfying the following property: With constant probability, for all $C \subset \mathbb{R}^{\tilde{D}}$ with $|C| = K$ and with

$$\phi_{\pi(S)}^{(m)}(C) \leq \alpha \cdot \phi_{\pi(S), K, m}^{opt}, \quad (31)$$

we have

$$\phi_{X}^{(m)}(C') \leq \alpha \cdot (1 + \epsilon) \phi_{X, K, m}^{OPT},$$

where $C'$ is the set of means induced with respect to $S$ by the responsibilities that are induced with respect to $\pi(S)$ by $C$. The algorithms’ runtime is bounded by

$$\mathcal{O}\left(N \log(\log(N))^3 \cdot D^3 \cdot K^{4m+1} \cdot m^3 \cdot \epsilon^{-10}\right).$$

Proof. Let $S$ be the set computed as in Theorem 7 given $X$, $K$, $\tilde{\epsilon} := \epsilon/4$, and some sufficiently small constant $\delta > 0$. Let $\pi$ be the dimension-reduction computed as in Lemma 10 with respect to $S$. Then, with constant probability, $S$ is a coreset for balanced solutions and Lemma 10 holds true as well.

To keep our notation uncluttered, we let $\mathcal{R}_Y(M)$ be the responsibilities that are induced by some set of means $M$ with respect to data set $Y$ and let $\mathcal{C}_Y(R)$ be the means induced by some set of responsibilities $R$ with respect to the data set $Y$.

Let $C \subset \mathbb{R}^D$ with $|C| = K$ and satisfying Equation 31. Due to the optimality of induced solutions, we have

$$\phi_{\pi(S)}^{(m)}(C) = \phi_{\pi(S)}^{(m)}(C, \mathcal{R}_{\pi(S)}(C)) \geq \phi_{\pi(S)}^{(m)}(\mathcal{R}_{\pi(S)}(C)). \quad (32)$$
By combining Equation (32) with Equation (31), we obtain
\[
\phi^{(m)}_{\pi(S)}(R_{\pi(S)}(C)) \leq \phi_{\pi(S)}(C) \leq \alpha \phi^{opt}_{(\pi(S),K,m)}.
\]
Due to Lemma [10], this means that
\[
\phi^{(m)}_{S}(R_{\pi(S)}(C)) \leq (1 + \tilde{\epsilon}) \phi^{opt}_{(S,K,m)} \tag{33}
\]
Due to the optimality of induced solutions, we have
\[
\phi^{(m)}_{S}(R_{\pi(S)}(C)) = \phi^{(m)}_{S}(C), R_{\pi(S)}(C) \geq \phi^{opt}_{S}(R_{\pi(S)}(C)). \tag{34}
\]
By combining Equation (33) and (34), we obtain
\[
\phi^{(m)}_{S}(R_{\pi(S)}(C)) \leq (1 + \tilde{\epsilon}) \phi^{opt}_{(S,K,m)}. \tag{35}
\]
With Theorem [8], we can conclude
\[
\phi^{(m)}_{X}(C), R_{\pi(S)}(C)) \leq \alpha(1 + \tilde{\epsilon})^{2} \phi^{opt}_{(X,K,m)} \\
\leq \alpha(1 + \epsilon) \phi^{opt}_{(X,K,m)}. \tag{36}
\]
(\(\tilde{\epsilon} = \epsilon/4\) and Lemma [16])

Proof of Corollary [11] Combining Theorem [9] with Theorem [32], we directly obtain Corollary [11].

References

[Aggarwal et al., 2009] Aggarwal, A., Deshpande, A., and Kannan, R. (2009). Adaptive sampling for k-means clustering. In Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, volume 5687 of Lecture Notes in Computer Science, pages 15–28. Springer Berlin Heidelberg.

[Bezdek et al., 1984] Bezdek, J., Ehrlich, R., and Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. Computers & Geosciences, 10(2):191–203.

[Bezdek et al., 1987] Bezdek, J., Hathaway, R., Sabin, M., and Tucker, W. (1987). Convergence theory for fuzzy c-means: Counterexamples and repairs. Systems, Man and Cybernetics, IEEE Transactions on, 17(5):873–877.
[Blömer et al., 2015] Blömer, J., Brauer, S., and Bujna, K. (2015). Complexity and approximation of the fuzzy k-means problem. CoRR, abs/1512.05947.

[Chen, 2009] Chen, K. (2009). On Coresets for K-Median and K-Means Clustering in Metric and Euclidean Spaces and Their Applications. SIAM J. Comput., 39(3):923–947.

[Dembélé and Kastner, 2003] Dembélé, D. and Kastner, P. (2003). Fuzzy C-means method for clustering microarray data. Bioinformatics, 19(8):973–980.

[Dunn, 1973] Dunn, J. C. (1973). A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. Journal of Cybernetics, 3(3):32–57.

[Haussler, 1992] Haussler, D. (1992). Decision theoretic generalizations of the PAC model for neural net and other learning applications. Information and Computation, 100(1):78 – 150.

[Inaba et al., 1994] Inaba, M., Katoh, N., and Imai, H. (1994). Applications of Weighted Voronoi Diagrams and Randomization to Variance-based K-clustering. In Proceedings of the Tenth Annual Symposium on Computational Geometry, SoCG ’94, pages 332–339, New York, NY, USA. ACM.

[Rezaee et al., 2000] Rezaee, M., van der Zwet, P., Lelieveldt, B., van der Geest, R., and Reiber, J. (2000). A multiresolution image segmentation technique based on pyramidal segmentation and fuzzy clustering. Image Processing, IEEE Transactions on, 9(7):1238–1248.