On Euclidean k-Means Clustering with α-Center Proximity

Amit Deshpande
Microsoft Research, India
amitdesh@microsoft.com

Anand Louis
Indian Institute of Science
anandl@iisc.ac.in

Apoorv Singh
Indian Institute of Science
apoorvsingh@iisc.ac.in

Abstract

k-means clustering is NP-hard in the worst case but previous work has shown efficient algorithms assuming the optimal k-means clusters are *stable* under additive or multiplicative perturbation of data. This has two caveats. First, we do not know how to efficiently verify this property of optimal solutions that are NP-hard to compute in the first place. Second, the stability assumptions required for polynomial time k-means algorithms are often unreasonable when compared to the ground-truth clusters in real-world data. A consequence of multiplicative perturbation resilience is *center proximity*, that is, every point is closer to the center of its own cluster than the center of any other cluster, by some multiplicative factor $\alpha > 1$.

We study the problem of minimizing the Euclidean k-means objective only over clusterings that satisfy α-center proximity. We give a simple algorithm to find the optimal α-center-proximal k-means clustering in running time exponential in k and $1/(\alpha - 1)$ but linear in the number of points and the dimension. We define an analogous α-center proximity condition for outliers, and give similar algorithmic guarantees for k-means with outliers and α-center proximity. On the hardness side we show that for any $\alpha' > 1$, there exists an $\alpha \leqslant \alpha'$, ($\alpha > 1$), and an $\varepsilon_0 > 0$ such that minimizing the k-means objective over clusterings that satisfy α-center proximity is NP-hard to approximate within a multiplicative $(1 + \varepsilon_0)$ factor.

1 Introduction

Clustering is an important tool in any data science toolkit. Most popular clustering algorithms partition the given data into disjoint clusters by optimizing a certain global objective such as the k-means. The implicit assumption in doing so is that an optimal solution for this objective would recover the underlying *ground truth* clustering. However, many such objectives are NP-hard to optimize in the worst case, e.g., k-center, k-median, k-means. Moreover, an optimal solution need not satisfy certain properties desired from the ground truth clustering, e.g., balance, stability. We highlight this problem using the example of k-means.

Given a set of n points $X = \{x_1, x_2, \ldots, x_n\}$ in a metric space with the underlying metric $\text{dist}(\cdot, \cdot)$, and a positive integer k, the k-means objective is to find centers $\mu_1, \mu_2, \ldots, \mu_k$ in the given metric space so as to minimize the sum of squared distances of all the points to their nearest centers, respectively, i.e., minimize $\sum_{i=1}^{n} \min_{1 \leq j \leq k} \text{dist}(x_i, \mu_j)^2$. This results in clusters $C_1, C_2, \ldots, C_k$, where the cluster C_j consists of all the points x_i whose nearest center is μ_j. Optimization of the k-means objective is NP-hard in the worst case, even when for Euclidean k-means with $k = 2$ [ADHP09, DF09] or $d = 2$ [MNV12]. There are known worst-case instances where the popular Lloyd's algorithm for k-means takes exponentially many iterations for convergence to a local optimum [Vat11]. On the algorithmic side, k-means++ initialization gives $O(\log k)$-approximation, and there are known $O(1)$-approximations in time $\text{poly}(n, k, d)$ [JV01] and $(1 + \varepsilon)$-approximations for the Euclidean k-means in time $O(2^{\text{poly}(k/\varepsilon)} \cdot nd)$ [KSS04, Che09]. Euclidean k-means is known to be NP-hard to approximate within some fixed constant $c > 1$ [ACKS15], hence, the exponential dependence on k is necessary for $(1 + \varepsilon)$-approximation.

In practice, however, Lloyd's algorithm, k-means++, and their variants perform well on most real-world data sets. This dichotomy between theoretical intractability and empirically observed efficiency has lead to the CDNM thesis [Ben15]: Clustering is difficult only when it does not matter! In most real-world data sets, the underlying ground-truth clustering is unambiguous and stable under small perturbations of data. As a consequence, the ground-truth clustering satisfies *center proximity*, that is, every point is closer to the center of its own cluster than the center of any other cluster, by some multiplicative factor $\alpha > 1$. Thus, center proximity is a desirable property for the output of any clustering algorithm used in practice. Balanced clusters of size $\Omega(n/k)$ is another desirable property to avoid small, meaningless clusters. Motivated by this, we study the problem of minimizing the k-means objective, where the minimization is only over clusterings that satisfy α-center proximity and are balanced.

Definition 1.1 (α-Center Proximity). Let $C_1, C_2, \ldots, C_k$ be a clustering of X with the centers $\mu_1, \mu_2, \ldots, \mu_k$ and the underlying metric $\text{dist}(\cdot, \cdot)$. We say that the clustering $C_1, C_2, \ldots, C_k$ of X satisfies α-center proximity if for all $i \neq j$ and $x \in C_i$, we have $\text{dist}(x, \mu_j) > \alpha \, \text{dist}(x, \mu_i)$. We say that a k-clustering is α-*center proximal* if it satisfies the α-center proximity property.

Unlike previous work, we do not assume that an optimal solution for the k-means objective on the given input satisfies α-center proximity. There is no easy way to algorithmically verify this promise. In fact, as we show later in this paper, there are instances where the optimal k-means solutions satisfy α-center proximity, for a small constant $\alpha > 1$, but the k-means problem still remains NP-hard. Therefore, we define our problem as finding a clustering of the smallest k-means cost among all the clusterings that satisfy α-center proximity.

Given any set of cluster centers, the outliers are the last few points when we order all the points in non-decreasing order of their distances to the nearest centers, respectively. For the set of outliers to be unambiguous and stable under small perturbations to the input data, intuitively one needs a multiplicative gap between the distances of inliers and outliers to their respective centers. So we define an analogous center proximity property for clustering with outliers as follows.

Definition 1.2 (α-Center Proximity with Outliers). Consider a k-means instance on the set of points X with underlying metric $\text{dist}(\cdot, \cdot)$, and an integer parameter z. We define the distance of a point x to a set or a tuple $(\mu_1, \ldots, \mu_n)$ as $\min_i \|x - \mu_i\|$. Given any centers $\mu_1, \mu_2, \ldots, \mu_k$, let $Z \subseteq X$ be the subset of the farthest z points in X based on their distances to the nearest center, and let $C_1, C_2, \ldots, C_k$ be the clustering of $X \setminus Z$ where C_j consists of the points in $X \setminus Z$ that have μ_j as their nearest center. Such a clustering $C_1, C_2, \ldots, C_k$ of $X \setminus Z$ with Z as outliers satisfies α-center proximity, if for all $i \neq j$ and $x \in C_i$, we have $\text{dist}(x, \mu_j) > \alpha \, \text{dist}(x, \mu_i)$, and moreover, for all $i, j \in [k]$, $x \in C_i$ and $y \in Z$ we have $\text{dist}(y, \mu_j) > \alpha \, \text{dist}(x, \mu_i)$.

1.1   Our results

We show that there exists a constant $c > 1$ such that k-means remains NP-hard to approximate within a multiplicative factor c, even on the instances where an optimal k-means solution satisfies α-center proximity and is balanced, i.e., each optimal cluster has size $\Omega(n/k)$. Moreover, for α close to 1, there may not be a unique optimal k-means solution that satisfies α-center proximity and is balanced. In fact, given any $\alpha > 1$, there exists an instance with $2^{\Omega(k/(\alpha-1))}$ such optimal k-means solutions that satisfy α-center proximity and are balanced.

For any $\alpha > 1$, we show an interesting geometric property (Proposition 2.3) for clusterings that satisfy α-center proximity, namely, any pair of disjoint clusters must lie inside two disjoint balls. The centers of these balls need not be at the means of the clusters, allowing the clusters to be arbitrarily large (see Figure 1). The degenerate case for $\alpha = 1$ is two balls of infinite radii touching at their separating hyperplane.

We show the following algorithmic result for minimizing the k-means objective over the clusterings that are balanced and satisfy α-center proximity.

Theorem 1.3. *For any $\alpha > 1$, a balance parameter $\omega > 0$, and given any set of n points in $\mathbb{R}^d$, we can exactly find a clustering of the least k-means cost among all solutions that satisfy α-center proximity and are balanced, i.e., each cluster has size at least $\omega n/k$. Our algorithm finds such an optimal clustering in time $O(2^{poly(k/\omega(\alpha-1))} nd)$, with constant probability.*

Remark 1.4. We remark that our algorithm requires α as an input. However, in practice, the value of α might not be available in general. For an input α, our algorithm can also be used to check whether an instance has an α-center proximal clustering. On invoking our algorithm with a certain value of α, if the instance has an α-center proximal clustering, then our algorithm will output the optimal α-center proximal clustering with constant probability. Therefore, the user can invoke our algorithm with sequence of decreasing values of α till a "satisfactory" clustering is found.

Since k-means is hard to approximate within some fixed constant $c > 1$, even on instances where the optimal solutions are balanced and satisfy α-center proximity, the exponential running time in our algorithm is unavoidable. We show the following hardness result:

Theorem 1.5. *For any $2 > \alpha' > 1$ there exists an $\alpha \leqslant \alpha'$, $(\alpha > 1)$, constants $\varepsilon > 0$, and $\omega > 0$, such that it is NP-hard to approximate the optimal α-center proximal Euclidean k-means, where the size of each cluster is at least $\omega n/k$, to a factor better than $(1 + \varepsilon)$.*

The running time of our algorithm is exponential only in the number of clusters k, the balance parameter ω and the center proximity parameter α but it is linear in the number of points n and the dimension d.

We show a similar exact algorithm for minimizing the k-means objective with z outliers, where the minimization is only over clusterings that satisfy the α-center proximity with outliers and are balanced.

Theorem 1.6. *For any $2 > \alpha > 1$, a balance parameter $\omega > 0$, given any set of n points in $\mathbb{R}^d$ and an outlier parameter $z \in [n]$, we can* exactly *find a clustering of the least k-means cost among all solutions that satisfy α-center proximity with z outliers and are balanced, i.e., each cluster has size at least $\omega n/k$. Our algorithm finds such an optimal clustering in time $O(2^{poly(k/\omega(\alpha-1))} nd)$, with constant probability.*

In the case when most points satisfy α-center proximity and form balanced clusters, we show an algorithm which outputs a list of clusterings, such that one of the clusterings corresponds to the case where the points which satisfy α-perturbation resilience are correctly clustered.

Theorem 1.7. *For any $\alpha > 1$, a balance parameter $\omega > 0$, given any set of n points in $\mathbb{R}^d$ and a parameter $\beta \in [n]$, we can output a list, of k-means clustering, of size $O\left(2^{poly(k/\omega(\alpha-1))}\right)$ such that one of them is the minimum cost clustering among all solutions that satisfy α-center proximity without β points and are balanced, i.e., each cluster has size at least $\omega n/k$. Our algorithm finds such an optimal clustering in time $O(2^{poly(k/\omega(\alpha-1))} nd)$, with constant probability.*

In fact, Theorem 1.3 and Theorem 1.6 hold for any clustering objective as long as the centers used to define α-center proximity are the *means* or *centroids* of the clusters.

We also show exact algorithm for minimizing the k-means objective over clustering that satisfy α-center proximity but no balance requirement. However, the running time of our algorithm depends exponentially on the ratio of the distances between the farthest and the closest pair of means.

$$\gamma^* = \frac{\max_{i,j} \|\mu_i - \mu_j\|}{\min_{i \neq j} \|\mu_i - \mu_j\|}.$$

Theorem 1.8. *For any $\alpha > 1$ and given any set of n points in $\mathbb{R}^d$, and a parameter γ, where*

$$\gamma \geq \frac{\max_{i,j} \|\mu_i - \mu_j\|}{\min_{i \neq j} \|\mu_i - \mu_j\|},$$

for the means $\mu_1, \mu_2, \ldots, \mu_k$ of the optimal solution, we can exactly *find a clustering of the least k-means cost among all solutions that satisfy α-center proximity. Our algorithm finds such an optimal clustering in time $O(2^{poly(k\gamma/(\alpha-1))} nd)$, with constant probability.*

We also show that the optimal α-center proximal clustering with balanced cluster need not be unique. In fact, the number of possible optimal α-center proximal clustering with balanced cluster can be exponential in k and $1/(\alpha - 1)$. We show the following result:

Proposition 1.9. *For any $2 > \alpha' > 1$, and any $k \in \mathbb{Z}$, there exists $\alpha \leq \alpha'$ (and $\alpha > 1$), n, d and a set of points $X \in \mathbb{R}^d$ such that such that the number of possible optimal α-center proximal clusterings, where the size of each cluster is n/k ($\omega = 1$) is $2^{\tilde{\Omega}(k/(\alpha^2-1))}$.*

1.2 Related Work

Metric Perturbation Resilience Bilu and Linial [BL12] introduced the notion of multiplicative perturbation resilience for discrete optimization problems. They showed results on Max-Cut problems, that if the instance is roughly $O(n)$ stable then they can retrieve the optimal Max-Cut in polynomial time. Later Makarychev et al. [MMV14] gave an algorithm which required only $(c\sqrt{\log n} \log\log n)$- multiplicative perturbation resilience for some constant $c > 0$ to retrieve the optimal Max-Cut in polynomial time. Bilu and Linial [BL12] had conjectured that there is some constant γ^* such that, γ^*-perturbation resilient instances can be solved in polynomial time. They had also asked the question if it can be proved that Max-Cut is NP-hard even for γ-perturbation resilient instances, for some constant γ. Awasthi et al. [ABS12] showed that their conjecture in the case of "well-behaved" center based objective functions like k-means, and k-median is true, by giving a polynomial time algorithm for 3-perturbation resilient instances. They proposed the definition of center-based clustering objective and α- center proximity (weaker notion than α-multiplicative perturbation resilience). They showed that solving k-median instances with Steiner points over general metrics that satisfy α-center proximity is NP-hard for $\alpha < 3$. Ben-David and Reyzin [BR14] showed that for every $\varepsilon > 0$, k-median instances with no Steiner points that satisfy $(2 - \varepsilon)$-center proximity are NP-hard. Balcan et al. [BHW16] showed that symmetric k-center under $(2 - \varepsilon)$-multiplicative perturbation resilience is hard unless NP=RP. They also show an algorithm for solving symmetric and asymmetric k-center for $\alpha \geq 2$. Balcan and Liang [BL16] improved upon the results of [ABS12] and showed that center based objective can be optimally clustered for $\alpha \geq 1 + \sqrt{2}$ factor perturbation resilience. They showed in their Lemma 3.3 that for a pair of cluster optimal clusters C_i, C_j, and $\alpha \geq 1 + \sqrt{2}$, the clusters are contained in disjoint balls around their centers μ_i and μ_j respectively. We show in Proposition 2.2 that for any $\alpha > 1$, and for a pair of optimal clusters C_i, C_j, the points in the clusters are contained in disjoint balls, centered around a point different from the mean of the clusters. We note that a similar geometric property was observed by Telgarsky and Vattani [TV10]. Telgarsky and Vattani [TV10] defined a version of Harting's method, which does updates based on the value α. The value of α in their case is defined in terms of the size of the clusters. Although in a different context, the geometric insight they achieved is the same as in Proposition 2.2. Recently Angelidakis et al. [AMM17] gave a more general definition of center based objective functions and of metric perturbation resilience compared to the one given by [ABS12]. They improved the previous results [BL16], giving an algorithm for center based clustering under 2-multiplicative perturbation resilience.

Independent and concurrent to this work, Friggstad et al. [FKS18] showed that for any fixed $d \geq 1$ and $\alpha > 1$, α-multiplicative perturbation resilient instances of discrete k-means and k-median (where the centers must be from

the data-points themselves) in metrics with doubling dimension d can be solved in time $O\left(n^{d^{O(d)}(\alpha-1)^{-O(d/(\alpha-1))}}k\right)$. They also showed that when the dimension d is a part of the input, there is a fixed $\varepsilon_0 > 0$ such there is not even a PTAS for $(1 + \varepsilon_0)$-multiplicative perturbation resilient instances of k-means in $\mathbb{R}^d$ unless NP=RP. We note that our hardness result does not subsume their hardness of approximation result as multiplicative perturbation resilient instances are a subset of center-proximal instances (that is, the optimal k-means solution satisfies center-proximity [AMM17]). Our problem is different from theirs; we look for center proximal instances which is a more general class of instances than perturbation resilient instances [AMM17]. Moreover, we do not assume that the optimal solution k-means solution is center proximal, unlike Friggstad et al. [FKS18], where they assume that the optimal k-means instance is perturbation resilient. Note that it is not known (to the best of our knowledge) whether there is an efficient algorithm to check whether an instance satisfies α-metric perturbation resilience.

Approximation Algorithms for Euclidean k-means   Kanungo et al. [KMN+04] proposed a $(9 + \varepsilon)$-approximation algorithm with running time $O\left(n^3\varepsilon^{-d}\right)$. Arthur and Vassilvitskii [AV07] showed an approximation ratio of $O(\log k)$, with running time $O(ndk)$, much superior to [KMN+04]. Recently Ahmadian et al. [ANFSW17] improved the approximation ratio given by [KMN+04] to $(6.357 + \varepsilon)$. There have been works to get a PTAS for Euclidean k-means objective. In order to obtain the PTAS, many have focused on cases where k or d or both are assumed to be fixed. Inaba et al. [IKI94] gave a PTAS when both k and d is fixed. There have been a series of work in the case when only k is assumed to be fixed [VKKR03, HPM04, HPK05, FMS07, KSS04, Che09]. Recently there has been works which give a PTAS for Euclidean k-means where only d is assumed to be a constant [CAKM16, FRS16].

Sampling Based Methods   As mentioned before that our results use sampling techniques often used in $(1 + \varepsilon)$-approximation for k-means [KSS04, Che09, ABS10]. The first ever linear (in n and d) running time for obtaining PTAS (assuming k to be a constant) given by [KSS04] is $O\left(nd2^{poly(k/\varepsilon)}\right)$. Feldman et al. [FMS07] gave a new algorithm (using efficient coreset construction) with a better running time than that of [KSS04] from $O\left(nd2^{poly(k/\varepsilon)}\right)$ to $O\left(nkd + d.poly(k/\varepsilon) + 2^{\tilde{O}(k/\varepsilon)}\right)$. There have been other works which also show similar results using D^2 sampling method [JKS14, BJK18]. Ding and Xu [DX15] gave a sampling based procedure to cluster other variants of k-means objective, which they called the constrained k-means clustering. These clustering objectives need not satisfy the locality property in Euclidean space. Their algorithm is based on uniform sampling and some stand alone geometric technique which they call the 'simplex lemma'. The running time of their algorithm is $O\left(2^{poly(k/\varepsilon)}n(\log n)^{k+1}d\right)$. Bhattacharya et al. [BJK18] later gave a more efficient algorithm based on D^2 sampling for the same class of constrained k-means problem and gave an algorithm with running time $O\left(2^{\tilde{O}(k/\varepsilon)}nd\right)$. Deshpande et al.[DLS18] also use the D^2 sampling method to solve the min-max k-means problem, wherein, one has to find a clustering such that the maximum cost of the cluster is minimized. All the work related to sampling based methods above estimate the means/centers of the clusters, and use them to recover a clustering.

Clustering with Outliers   For the problem of k-median with outliers, Charikar et al. [CKMN01] gave an algorithm that removes $(1 + \varepsilon)$ times the number of outliers, and has a cost at most $4(1 + 1/\varepsilon)$ times the optimal cost. Later, Chen [Che08] gave a $O(1)$ approximation for the k-median with outliers problem. However, the unspecified approximation factor of their algorithm is large. For the problem of k-means with outliers, Gupta et al. [GKL+17] gave a $O(1)$ approximation by removing $O(z\, k \log n)$ number of outliers (where z is the number of outliers). Friggstad et al. [FKRS18] gave an algorithms that open $k(1 + \varepsilon)$ facilities, and have an approximation factor of $(1 + \varepsilon)$ on Euclidean and doubling metrics, and $25 + \varepsilon$ on general metrics. Recently, Krishnaswamy et al. [KLS18] gave a $(53.002 + \varepsilon)$ factor approximation for the k-means with outliers problem, and a $(7.081 + \varepsilon)$ approximation factor for the k-median with outliers problem, without violating any constraint. Chekuri et al. [CG18] defined a new model of perturbation resilience for clustering with outliers, and exactly solve the clustering with outliers problem for several common center based objectives (like k-center, k-means, k-median) when the instances is 2- multiplicative perturbation resilient. We note that our definition of outliers is inspired by Chekuri et al. [CG18].

Other Notions of Stability   Ackerman and Ben-David [ABD09] studied various deterministic assumptions to obtain solutions with small objective costs, one of them being additive perturbation resilience. We refer the reader to [ABD09]

for a more detailed survey of various notions of stability and their implications. Vijayraghavan et al. [VDW17] gave a more general definition of ε-additive perturbation resilience, where the points in the instance even after being moved by at most $\varepsilon \cdot \max_{i,j} \|\mu_i - \mu_j\|$ (where $\varepsilon \in (0, 1)$ is a parameter) reamin in the same optimal clusters. Vijayaraghavan et al. [VDW17] showed a geometric property of ε-additive perturbation resilient instances implies an *angular separation* between points from any pair of clusters. Using this observation, they showed that a modification of the perceptron algorithm (from supervised learning) can optimally solve any ε-additive perturbation resilient instance of 2-means in time $dn^{\text{poly}(1/\varepsilon)}$. We make an observation (Proposition 2.3) that α-metric perturbation resilient instances of k-means in the Euclidean space also satisfy *angular separation* similar to [VDW17]. For k-means their running time is dn^{k^2/ε^8}. To get a faster algorithm, they defined a stronger notion of stability than ε-additive perturbation resilience called $(\rho, \Delta, \varepsilon)$-separation, a natural strengthening of additive perturbation stability where there is an additional margin of ρ between any pair of clusters. They give an algorithm based on the k-largest connected components in a graph that can optimal solve any $(\rho, \Delta, \varepsilon)$-separated instance of k-means with β-balanced clusters in time $\tilde{O}(n^2 kd)$ whenever $\rho = \Omega(\Delta/\varepsilon^2 + \beta\Delta/\varepsilon)$. They also showed that their algorithm is robust to outliers as long as the fraction η of outliers satisfies the following equation

$$\rho = \Omega\left(\frac{\Delta}{\varepsilon^2}\left(\frac{w_{\max} + \eta}{w_{\min} - \eta}\right)\right),$$

where $w_{\max}$ and $w_{\min}$ are the fraction of points in the largest and the smallest optimal cluster, respectively. Balcan et al. [BBG13] explored the (c, ε)-approximation stability which assumes that every c-approximation to the cost is ε-close (in normalized set difference) to the target clustering . Balcan and Liang [BL16] showed that when the target clustering is the optimal clustering then the (c, ε)-approximation stability implies (c, ε)-perturbation resilience, that is, the optimum after perturbation of up to a multiplicative factor c is ε-close (in normalized set difference) to the original clustering. Kumar and Kannan [KK10] also gave deterministic conditions (*read stability*) under which their algorithm finds the optimal clusters. The analysis of Kumar and Kannan [KK10] was tightened by [AS12] which basically needed that the cluster centers must be pair-wise separated by a margin of $\Omega(\sqrt{k}\sigma)$ along the line joining the mean of the clusters, where σ denotes the "spectral radius" of the data-set.

1.3   Notation

We use $d \in \mathbb{Z}$ to denote the dimension of the ambient space. For a vector $v \in \mathbb{R}^d$, we use $\|v\|$ to denote its Euclidean norm. We use $x_1, \ldots, x_n \in \mathbb{R}^d$ to denote the points in the instance, and we use X to denote the set of these points. For any set of points S, we define its *diameter* as $\text{diam}(S) \overset{\text{def}}{=} \max_{u,v \in S} \|u - v\|$.

2   Geometric Properties of α-Center Proximal Instances

We will assume Euclidean metric throughout the paper.

Definition 2.1. Suppose there exists a clustering $C_1, \ldots, C_k$ of X that is α-center proximal. Let C_i and C_j be two clusters with μ_i and μ_j as their respective means. We define the vector $\hat{\mu}_{i,j} \overset{\text{def}}{=} \frac{\alpha^2 \mu_i - \mu_j}{\alpha^2 - 1}$ and $r_{i,j} \overset{\text{def}}{=} \frac{\alpha}{\alpha^2 - 1}\|\mu_i - \mu_j\|$. Let $C_{i,j}$ denote the ball centered at $\hat{\mu}_{i,j}$ with radius $r_{i,j}$. Let $p_{i,j} \overset{\text{def}}{=} (\hat{\mu}_{i,j} + \hat{\mu}_{j,i})/2$ and $D_{i,j} \overset{\text{def}}{=} \|\hat{\mu}_{i,j} - \hat{\mu}_{j,i}\|$. Let $u \overset{\text{def}}{=} \frac{(\hat{\mu}_{i,j} - \hat{\mu}_{j,i})}{\|\hat{\mu}_{i,j} - \hat{\mu}_{j,i}\|}$ denote the unit vector along the line joining $\hat{\mu}_{i,j}$ and $\hat{\mu}_{j,i}$, and let $d_{i,j}$ denote the distance between the closest points in $C_{i,j}$ and $C_{j,i}$.

We note that Balcan and Liang [BL16] show in their Lemma 3.3 that for a pair of clusters C_i, C_j in a α-center proximal clustering, and $\alpha \geqslant 1 + \sqrt{2}$, the clusters are contained in disjoint balls around their centers μ_i and μ_j respectively. We show in the following proposition that for any $\alpha > 1$, and for a pair of clusters C_i, C_j in a α-center proximal clustering, the points in the clusters are contained in disjoint balls, centered around $\hat{\mu}_{i,j}$ and $\hat{\mu}_{j,i}$.

Proposition 2.2 (Geometric implication of α-center proximity property)**.** *Let X satisfy the α-center proximity property (for any $\alpha > 1$) and let C_i and C_j be two clusters in its optimal solution. Any point $x \in C_i$, satisfies*

$$\forall x \in C_i, \|x - \hat{\mu}_{i,j}\| < r_{i,j}. \tag{1}$$

5

Proof. The proof proceeds simply by using the α-center proximity property for Euclidean metric and squaring both the sides, we get

$$\alpha^2 \|x - \mu_i\|^2 < \|x - \mu_j\|^2 \, ,$$

$$\alpha^2 \|x\|^2 + \alpha^2 \|\mu_i\|^2 - 2\alpha^2 \langle x, \mu_i \rangle < \|x\|^2 + \|\mu_j\|^2 - 2 \langle x, \mu_j \rangle \, ,$$

$$\|x\|^2 - 2 \left\langle x, \frac{\alpha^2 \mu_i - \mu_j}{(\alpha^2 - 1)} \right\rangle < \frac{\|\mu_j\|^2 - \alpha^2 \|\mu_i\|^2}{(\alpha^2 - 1)} \, .$$

Completing the square on LHS and adding the appropriate term on RHS

$$\|x\|^2 - 2 \left\langle x, \frac{\alpha^2 \mu_i - \mu_j}{(\alpha^2 - 1)} \right\rangle + \left\| \frac{\alpha^2 \mu_i - \mu_j}{(\alpha^2 - 1)} \right\|^2 < \frac{\|\mu_j\|^2 - \alpha^2 \|\mu_i\|^2}{(\alpha^2 - 1)} + \left\| \frac{\alpha^2 \mu_i - \mu_j}{(\alpha^2 - 1)} \right\|^2 \, ,$$

$$\left\| x - \frac{\alpha^2 \mu_i - \mu_j}{(\alpha^2 - 1)} \right\|^2 < \frac{\alpha^2 \|\mu_i - \mu_j\|^2}{(\alpha^2 - 1)^2} \, .$$

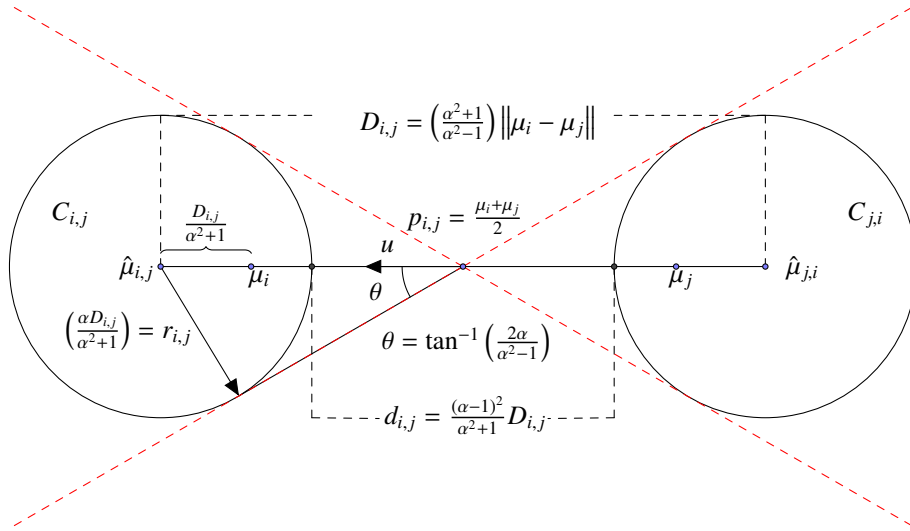Therefore using the terms in Definition 2.1 we get the proposition.

$\square$

Figure 1: Geometric implication of α-center proximity property.

Proposition 2.3. *Suppose there exists a clustering $C_1, ..., C_k$ of X that is α-center proximal (for any $\alpha > 1$). From Proposition 2.2 we know that for any two clusters C_i and C_j, the clusters $(i \neq j)$ lie inside disjoint balls centered at $\hat{\mu}_{i,j}$ and $\hat{\mu}_{j,i}$ with radius r_{ij} and r_{ji}, respectively. The following structural properties for α-center proximal clustering holds:*

(a) *Distance between the respective centers of the balls is $D_{i,j} = \left\| \hat{\mu}_{i,j} - \hat{\mu}_{j,i} \right\| = \frac{\alpha^2+1}{\alpha^2-1} \left\| \mu_i - \mu_j \right\|$.*

(b) *The mid-point of the line joining the respective centers of the two balls is $p_{i,j} = \frac{\hat{\mu}_{i,j} + \hat{\mu}_{j,i}}{2} = \frac{\mu_i + \mu_j}{2}$.*

(c) *The distance between the closest points in the two balls is greater than $d_{i,j} = \frac{\alpha-1}{\alpha+1} \left\| \mu_i - \mu_j \right\| = \frac{(\alpha-1)^2}{\alpha^2+1} D_{i,j}$.*

6

(d) *The radius of the ball $C_{i,j}$ is $r_{i,j} = \frac{\alpha}{\alpha^2-1} \left\| \mu_i - \mu_j \right\| = \frac{\alpha}{(\alpha-1)^2} d_{i,j} = \frac{\alpha}{\alpha^2+1} D_{i,j}$.*

(e) *Distance between the mean of the cluster and the center of the ball corresponding to the cluster is $\left\| \hat{\mu}_{i,j} - \mu_i \right\| = \frac{1}{\alpha^2-1} \left\| \mu_i - \mu_j \right\| = \frac{1}{(\alpha-1)^2} d_{i,j} = \frac{1}{\alpha^2+1} D_{i,j}$.*

(f) *The two clusters lie inside a cone with apex at $p_{i,j}$ whose axis is along the line joining $(\hat{\mu}_{i,j} - \hat{\mu}_{j,i})$ with half-angle $\tan^{-1}\left(\frac{2\alpha}{\alpha^2-1} \right)$.*

(g) *The diameter of the cluster C_i can be bounded by the radius of the ball $C_{i,j}$, $\operatorname{diam}(C_i) \leqslant \frac{2\alpha}{\alpha^2-1} \left\| \mu_i - \mu_j \right\| = \frac{2\alpha}{\alpha^2+1} D_{i,j}$.*

(h) *For any $x \in C_i, \left\| x - \mu_j \right\| > \frac{\alpha}{\alpha+1} \left\| \mu_i - \mu_j \right\|$.*

Proof. (a) $D_{i,j} = \left\| \hat{\mu}_{i,j} - \hat{\mu}_{j,i} \right\|$.

$$D_{i,j} = \left\| \frac{\alpha^2 \mu_i - \mu_j}{\alpha^2 - 1} - \frac{\alpha^2 \mu_j - \mu_i}{\alpha^2 - 1} \right\| = \left(\frac{\alpha^2 + 1}{\alpha^2 - 1} \right) \left\| \mu_i - \mu_j \right\| .$$

(b)

$$p_{i,j} = \frac{\hat{\mu}_{i,j} + \hat{\mu}_{j,i}}{2} = \frac{1}{2} \left(\frac{\alpha^2 \mu_i - \mu_j}{\alpha^2 - 1} + \frac{\alpha^2 \mu_j - \mu_i}{\alpha^2 - 1} \right) = \frac{\mu_i + \mu_j}{2} .$$

Note that $p_{i,j}$ is also the mid-point of the line joining the means of the two clusters.

(c) The distance between the closest points in the two balls is greater than $d_{i,j} = D_{i,j} - 2r_{i,j}$. Putting in the value of $D_{i,j}$ from part (a) of the proposition, and $r_{i,j}$ from the previous proposition, we get that

$$d_{i,j} > \left(\frac{\alpha^2 + 1}{\alpha^2 - 1} - \frac{2\alpha}{\alpha^2 - 1} \right) \left\| \mu_i - \mu_j \right\| = \frac{(\alpha-1)^2}{\alpha^2 - 1} \left\| \mu_i - \mu_j \right\| = \frac{(\alpha-1)^2}{\alpha^2 + 1} D_{i,j} .$$

(d) The proof follows from the definition of $r_{i,j}$ and part (c) of the proposition.

(e)

$$\left\| \hat{\mu}_{i,j} - \mu_i \right\| = \left\| \frac{\alpha^2 \mu_i - \mu_j}{\alpha^2 - 1} - \mu_i \right\| = \frac{1}{\alpha^2 - 1} \left\| \mu_i - \mu_j \right\| .$$

The proof follows using the part (c) of the proposition.

(f) Let the apex of the cone (with its axis along u) be at $p_{i,j}$. Draw the tangent from $p_{i,j}$ to $C_{i,j}$. Using the fact that the tangent is perpendicular to radius drawn on the point of contact of the ball and the tangent line, we get for the half-angle θ that $\sin \theta = \frac{r_{i,j}}{D_{i,j}/2} = \frac{\alpha/(\alpha^2+1)}{1/2}$. Therefore $\tan \theta = \frac{2\alpha}{\alpha^2-1}$.

(g) The equation 1 implies that all the points belonging to the cluster C_i must lie inside the ball $C_{i,j}$ of radius $r_{i,j}$. Therefore the proof follows using part (d) of the proposition.

(h) The smallest distance between any point in C_i and μ_j is greater than $r_{i,j} + d_{i,j} - \left\| \hat{\mu}_{j,i} - \mu_j \right\|$. Using part (d) and (e) of the proposition, we get

$$r_{i,j} + d_{i,j} - \left\| \hat{\mu}_{j,i} - \mu_j \right\| = \left(\frac{\alpha}{\alpha^2 - 1} + \frac{(\alpha-1)^2}{\alpha^2 - 1} - \frac{1}{\alpha^2 - 1} \right) \left\| \mu_i - \mu_j \right\| = \frac{\alpha}{\alpha + 1} \left\| \mu_i - \mu_j \right\| .$$

$\square$

Remark 2.4. The line joining the respective centers of the ball passes through the mean of the clusters as well. Refer to the Figure 1 for getting an insight into the geometric structure defined by the equation 1. We refer the reader to figure 1a of [VDW17] for insight into the geometry of ε-additive perturbation resilient instance, and to notice the similarity between the geometric structures of instances satisfying the two stability properties.

Algorithm 1: α-center proximal k-means clustering with balanced clusters

Input: a list $\mathcal{L}$ of k-tuples, where there exists at least one tuple $(\tilde{\mu}_1, \ldots, \tilde{\mu}_k)$ such that for any $i \in [k]$,
$\|\tilde{\mu}_i - \mu_i\| \leqslant 2\delta r_{i,j}$ for all $j \in [k]$, a number $\alpha > 1$, $\omega > 0$

Output: An α-center proximal clustering of minimum cost where the size of each cluster is at least $\omega n/k$.

1: Initialize $(\tilde{\mu}_1^{(p)}, \tilde{\mu}_2^{(p)}, \ldots, \tilde{\mu}_k^{(p)}) \leftarrow (\bar{0}, \bar{0}, \ldots, \bar{0})$

2: $\text{cost} \leftarrow \sum_{i=1}^{n} \min_{1 \leqslant j \leqslant k} \left\| x_i - \tilde{\mu}_j^{(0)} \right\|^2$

3: **for** a tuple $\tilde{\mu}^{(q)} \in \mathcal{L}$ **do**

4:      **for** $j = 1$ to k **do**

5:         $C_j^{(q)} = \{x_i \ : \ \tilde{\mu}_j^{(q)}$ is the nearest center to $x_i\}$

6:      **end for**

7:      **if** $\sum_{i=1}^{n} \min_{1 \leqslant j \leqslant k} \left\| x_i - \tilde{\mu}_j^{(q)} \right\|^2 < \text{cost}$ **then**

8:         **if** $C_1^{(q)}, \ldots, C_k^{(q)}$ all have size at least $\frac{\omega n}{k}$ and satisfy α-center proximity **then**

9:            $(\tilde{\mu}_1^{(p)}, \tilde{\mu}_2^{(p)}, \ldots, \tilde{\mu}_k^{(p)}) \leftarrow (\tilde{\mu}_1^{(q)}, \tilde{\mu}_2^{(q)}, \ldots, \tilde{\mu}_k^{(q)})$

10:            $\text{cost} = \sum_{i=1}^{n} \min_{1 \leqslant j \leqslant k} \left\| x_i - \tilde{\mu}_j^{(q)} \right\|^2$

11:         **end if**

12:      **end if**

13: **end for**

14: **for** $i = 1$ to n **do**

15:      $\text{Label}(i) \leftarrow \underset{1 \leqslant j \leqslant k}{\arg\min} \left\| x_i - \tilde{\mu}_j^{(p)} \right\|$

16: **end for**

2.1 Estimating Means Suffices for Cluster Recovery

In this section we assume that we can get access to a set of points $\{\tilde{\mu}_1, \ldots, \tilde{\mu}_k\}$ such that, $\|\tilde{\mu}_i - \mu_i\| \leqslant 2\delta r_{i,j}$ for all $j \in [k]$ and for each $i \in [k]$, with probability some constant probability. We will show how to obtain such points in Section 3.

The following proposition shows that to cluster a point to their desired cluster, an additive approximation to the center also works (deciding based on the proximity of the data points to the approximate center).

Proposition 2.5. *Suppose we have a set of points $\{\tilde{\mu}_1, \ldots, \tilde{\mu}_k\}$ such that for any $i \in [k]$, $\|\tilde{\mu}_i - \mu_i\| \leqslant 2\delta r_{i,j}$ for all $j \in [k]$, then we can find the optimal clustering $C_1, \ldots, C_k$.*

Proof. We will break the proof of the proposition into the following lemmas.

Lemma 2.6. *For any $j \in [k]$, for all $i \in [k]$, and for any point x, we have*

$$\left| \|x - \tilde{\mu}_j\| - \|x - \mu_j\| \right| \leqslant 2\delta r_{i,j} \, .$$

Proof. For any two vectors $u, v \in \mathbb{R}^d$, the triangle inequality implies that $\left| \|u\| - \|v\| \right| \leqslant \|u - v\|$. Therefore,

$$\left| \|x - \tilde{\mu}_j\| - \|x - \mu_j\| \right| \leqslant \|\tilde{\mu}_j - \mu_j\| \leqslant 2\delta r_{i,j} \, .$$

$\square$

Lemma 2.7. *Fix any $i \in [k]$. For any point $x \in C_i$, we have*

$$\|x - \tilde{\mu}_i\| < \|x - \tilde{\mu}_j\| \qquad \text{for each } j \in [k] \setminus \{i\} \, .$$

Proof. By our choice of parameters in Algorithm 4, we have

$$\delta \leqslant \frac{(\alpha - 1)^2}{8\alpha} \, . \tag{2}$$

8

Algorithm 2: α-center proximal k-means clustering where the distance between the means is bounded

---

Input: a list $\mathcal{L}$ of k-tuples, where there exists at least one tuple $(\tilde{\mu}_1, \ldots, \tilde{\mu}_k)$ such that for any $i \in [k]$,
$\|\tilde{\mu}_i - \mu_i\| \leqslant 2\delta r_{i,j}$ for all $j \in [k]$, a number $\alpha > 1$, $\gamma \geqslant 1$

Output: An α-center proximal clustering of minimum cost where $\gamma \leqslant \dfrac{\max_{i,j} \|\mu_i - \mu_j\|}{\min_{i \neq j} \|\mu_i - \mu_j\|}$

1: Initialize $(\tilde{\mu}_1^{(p)}, \tilde{\mu}_2^{(p)}, \ldots, \tilde{\mu}_k^{(p)}) \leftarrow (\bar{0}, \bar{0}, \ldots, \bar{0})$

2: cost $\leftarrow \sum_{i=1}^{n} \min_{1 \leqslant j \leqslant k} \left\| x_i - \tilde{\mu}_j^{(0)} \right\|^2$

3: **for** a tuple $\tilde{\mu}^{(q)} \in \mathcal{L}$ **do**

4:     **for** $j = 1$ to k **do**

5:         $C_j^{(q)} = \{x_i \; : \; \tilde{\mu}_j^{(q)}$ is the nearest center to $x_i\}$

6:     **end for**

7:     Find the actual means $(\mu_1^{(q)}, \ldots, \mu_k^{(q)})$ of $C_1^{(q)}, \ldots, C_k^{(q)}$

8:     **if** $\sum_{i=1}^{n} \min_{1 \leqslant j \leqslant k} \left\| x_i - \tilde{\mu}_j^{(q)} \right\|^2 <$ cost **then**

9:         **if** $C_1^{(q)}, \ldots, C_k^{(q)}$ satisfy α-center proximity and $\gamma \geqslant \dfrac{\max_{i,j} \left\| \mu_i^{(p)} - \mu_j^{(p)} \right\|}{\min_{i \neq j} \left\| \mu_i^{(p)} - \mu_j^{(p)} \right\|}$ **then**

10:             $(\tilde{\mu}_1^{(p)}, \tilde{\mu}_2^{(p)}, \ldots, \tilde{\mu}_k^{(p)}) \leftarrow (\tilde{\mu}_1^{(q)}, \tilde{\mu}_2^{(q)}, \ldots, \tilde{\mu}_k^{(q)})$

11:             cost $= \sum_{i=1}^{n} \min_{1 \leqslant j \leqslant k} \left\| x_i - \tilde{\mu}_j^{(q)} \right\|^2$

12:         **end if**

13:     **end if**

14: **end for**

15: **for** $i = 1$ to n **do**

16:     Label$(i) \leftarrow \underset{1 \leqslant j \leqslant k}{\operatorname{argmin}} \left\| x_i - \tilde{\mu}_j^{(p)} \right\|$

17: **end for**

Fix any $j \in [k] \setminus \{i\}$. Using Definition 1.1, we get

$$\|x - \mu_i\| < \frac{1}{\alpha} \|x - \mu_j\| . \tag{3}$$

$$
\begin{aligned}
\|x - \tilde{\mu}_i\| &\leqslant \|x - \mu_i\| + 2\delta r_{i,j} && \text{(Lemma 2.6)}\\
&< \frac{1}{\alpha} \|x - \mu_j\| + 2\delta r_{i,j} && \text{(Using (3))}\\
&= \left(\|x - \mu_j\| - 2\delta r_{i,j}\right) + \left(\frac{1}{\alpha} - 1\right)\|x - \mu_j\|\\
&\quad + 2\delta r_{i,j} + 2\delta r_{i,j}\\
&\leqslant \|x - \tilde{\mu}_j\| + 4\delta r_{i,j} - \frac{\alpha - 1}{\alpha}(\alpha - 1) r_{i,j} && \text{(Using Lemma 2.6 and Proposition 2.3)}\\
&\leqslant \|x - \tilde{\mu}_j\| + r_{i,j}\left(4\delta - \frac{(\alpha - 1)^2}{\alpha}\right)\\
&\leqslant \|x - \tilde{\mu}_j\| + r_{i,j} \cdot \left(-\frac{(\alpha - 1)^2}{2\alpha}\right) < \|x - \tilde{\mu}_j\| . && \text{(Using (2))}
\end{aligned}
$$

□

The next lemma shows that the clustering obtained by approximate centers corresponds to the optimal clustering, ignoring the outliers.

Lemma 2.8. *For any* $i \in [k]$, $\tilde{C}_i \cap (X \setminus Z) = C_i$.

Proof. Lemma 2.7 implies that each $x \in C_i$ is closest to $\tilde{\mu}_i$ out of $\{\tilde{\mu}_1, \ldots, \tilde{\mu}_k\}$. Therefore, $x \in \tilde{C}_i$, and hence

$$C_i \subseteq \tilde{C}_i \cap (X \setminus Z) \qquad \text{for each } i \in [k] . \tag{4}$$

By construction, each $x \in X$ belongs to exactly one out of $\{\tilde{C}_1, \ldots, \tilde{C}_k\}$. Therefore,

$$\left(\tilde{C}_i \cap (X \setminus Z)\right) \cap \left(\tilde{C}_j \cap (X \setminus Z)\right) = \emptyset \qquad \text{for each } i, j \in [k], i \neq j , \tag{5}$$

and,

$$\cup_{i \in [k]} \left(\tilde{C}_i \cap (X \setminus Z)\right) = X \cap (X \setminus Z) = \cup_{i \in [k]} C_i . \tag{6}$$

(4), (5) and (6) imply the lemma. □

Therefore the Lemma 2.8 implies the statement of the proposition, as in this setting, since $X \setminus Z = X$, Lemma 2.8 implies that $\tilde{C}_i = C_i$. □

We are now ready to prove Theorem 1.3 and Theorem 1.8

Proof of Theorem 1.3. Using Proposition 3.1 for a desired clustering $C_1, \ldots, C_k$, we get a set of points $\{\tilde{\mu}_1, \ldots, \tilde{\mu}_k\}$, such that

$$\mathbb{P}\left[\|\tilde{\mu}_i - \mu_i\| \leqslant \delta \operatorname{diam}(C_i) \text{ for each } i \in [k]\right] \geqslant 1 - \beta_1 .$$

Using Proposition 2.2, we get that for each $i \in [k]$ $\operatorname{diam}(C_i) \leqslant 2r_{i,j}$ for all $j \in [k]$. Therefore this implies that

$$\mathbb{P}\left[\|\tilde{\mu}_i - \mu_i\| \leqslant 2\delta r_{i,j} \text{ for all } j \in [k], \text{ for each } i \in [k]\right] \geqslant 1 - \beta_1 .$$

This can be achieved by Algorithm 4

Now, using Proposition 2.5 we get that we can get an exact clustering. This step is achieved by Algorithm 1. □

Proof of Theorem 1.8. Using Proposition 3.5 for a desired clustering $C_1, \ldots, C_k$, we get a set of points $\{\tilde{\mu}_1, \ldots, \tilde{\mu}_k\}$, such that

$$\|\tilde{\mu}_i - \mu_i\| \leqslant 2\delta r_{i,j} \text{ for all } j \in [k] \text{ and for each } i \in [k] .$$

with constant probability. This step is achieved by Algorithm 5.

Now, using Proposition 2.5 we get that we can get an exact clustering. This step is achieved by Algorithm 2.

The running time of the algorithm is $O\left(2^{\mathsf{poly}(k/\varepsilon)}nd\right)$, and from equation (14) and (2) we get that

$$\varepsilon \leqslant \left(\frac{2(\alpha-1)^2}{8\alpha k\gamma}\right)^2 .$$

Therefore, we get a running time of $O\left(2^{\mathsf{poly}(k\gamma/(\alpha-1))}nd\right)$. $\qquad\square$

2.2 α-Outliers Center Proximity

In this section we assume that we can get access to a set of points $\{\tilde{\mu}_1, \ldots, \tilde{\mu}_k\}$ such that, $\|\tilde{\mu}_i - \mu_i\| \leqslant \delta \, \mathsf{diam}\,(C_i)$ for all $i \in [k]$, with probability some constant probability. We will show how to obtain such points in Section 3.1

Algorithm 3: α-center proximal k-means clustering with balanced clusters and outliers

Input: a list $\mathcal{L}$ of k-tuples, where there exists at least one tuple $(\tilde{\mu}_1, \ldots, \tilde{\mu}_k)$ such that for any $i \in [k]$,
$\quad\|\tilde{\mu}_i - \mu_i\| \leqslant \delta \, \mathsf{diam}\,(C_i)$ for all $j \in [k]$, a number $\alpha > 1$, $\omega > 0$, and $z > 0$

Output: An α-center proximal clustering where the size of each cluster is at least $\omega n/k$.

1: Initialize $(\tilde{\mu}_1^{(p)}, \tilde{\mu}_2^{(p)}, \ldots, \tilde{\mu}_k^{(p)}) \leftarrow (\bar{0}, \bar{0}, \ldots, \bar{0})$

2: $\text{cost} \leftarrow \sum_{i=1}^{n} \min_{1 \leqslant j \leqslant k} \left\|x_i - \tilde{\mu}_j^{(0)}\right\|^2$

3: **for** a tuple $\tilde{\mu}^{(q)} \in \mathcal{L}$ **do**

4: $\quad$ **for** $j = 1$ to k **do**

5: $\quad\quad$ $C_j^{(q)} = \{x_i \ : \ \tilde{\mu}_j^{(q)} \text{ is the nearest center to } x_i\}$

6: $\quad$ **end for**

7: $\quad$ Remove the z farthest points with respect to $(\tilde{\mu}_1^{(p)}, \tilde{\mu}_2^{(p)}, \ldots, \tilde{\mu}_k^{(p)})$

8: $\quad$ **if** $\sum_{i=1}^{n} \min_{1 \leqslant j \leqslant k} \left\|x_i - \tilde{\mu}_j^{(q)}\right\|^2 < \text{cost}$ **then**

9: $\quad\quad$ **if** $C_1^{(q)}, \ldots, C_k^{(q)}$ all have size at least $\frac{\omega n}{k}$ and satisfy α-center proximity **then**

10: $\quad\quad\quad$ $(\tilde{\mu}_1^{(p)}, \tilde{\mu}_2^{(p)}, \ldots, \tilde{\mu}_k^{(p)}) \leftarrow (\tilde{\mu}_1^{(q)}, \tilde{\mu}_2^{(q)}, \ldots, \tilde{\mu}_k^{(q)})$

11: $\quad\quad\quad$ $\text{cost} = \sum_{i=1}^{n} \min_{1 \leqslant j \leqslant k} \left\|x_i - \tilde{\mu}_j^{(q)}\right\|^2$

12: $\quad\quad$ **end if**

13: $\quad$ **end if**

14: **end for**

15: **for** $i = 1$ to n **do**

16: $\quad$ $\text{Label}(i) \leftarrow \underset{1 \leqslant j \leqslant k}{\arg\min} \left\|x_i - \tilde{\mu}_j^{(p)}\right\|$

17: **end for**

We first note that, by our Definition 1.2, outliers are the farthest most points in the data-set, where distance is measured from the set of optimal centers $\mu_1, \ldots, \mu_k$.

Lemma 2.9. *The Definition 1.2 implies that for any $q \in Z$*

$$\left\|q - \mu_j\right\| \geqslant \alpha \frac{\mathsf{diam}\,(C_i) + \mathsf{diam}\left(C_j\right)}{4}, \qquad (\text{for all } i, j \in [k]) , \tag{7}$$

11

Proof. From the Definition 1.2 we get that for any $p_i \in C_i$, $p_j \in C_j$ and $q \in Z$, we get that $\alpha \|p_i - \mu_i\| \leqslant \|q - \mu_j\|$ and $\alpha \|p_j - \mu_j\| \leqslant \|q - \mu_j\|$. There exists at least one point $p_{i_1} \in C_i$ such that $\|p_{i_1} - \mu_i\| \geqslant \mathsf{diam}\,(C_i)/2$, this is because the diameter is defined as $\max_{p_{i_1}, p_{i_2} \in C_i} \|p_{i_1} - pi_2\|$. We will prove this by contradiction. Suppose $\|p_1 - p_2\| = \mathsf{diam}\,(C_i)$, and $\|p_{i_1} - \mu_i\| < \mathsf{diam}\,(C_i)/2$, and $\|p_{i_2} - \mu_i\| < \mathsf{diam}\,(C_i)/2$, then this contradicts the triangle inequality, as $\|p_1 - p_2\| > \|p_{i_1} - \mu_i\| + \|p_{i_2} - \mu_i\|$.

Therefore we get that $\|q - \mu_j\| \geqslant \alpha\, \mathsf{diam}\,(C_i)/2$ and $\|q - \mu_j\| \geqslant \alpha\, \mathsf{diam}\left(C_j\right)/2$. Therefore the statement of the lemma follows. □

Lemma 2.10. *For any $q \in Z$ and $x \in C_i$, we have that $\|x - \tilde{\mu}_i\| < \|q - \tilde{\mu}_j\|$, for $i, j \in [k]$.*

Proof.

$$\|x - \tilde{\mu}_i\| \leqslant \|x - \mu_i\| + \delta\, \mathsf{diam}\,(C_i) \qquad\qquad \text{(Lemma 2.6)}$$

$$< \frac{1}{\alpha} \|q - \mu_j\| + \delta\, \mathsf{diam}\,(C_i) \qquad\qquad \text{(Definition 1.2)}$$

$$= \left(\|q - \mu_j\| - \delta\, \mathsf{diam}\left(C_j\right)\right) + \left(\frac{1}{\alpha} - 1\right)\|q - \mu_j\|$$

$$\qquad + \delta\, \mathsf{diam}\,(C_i) + \delta\, \mathsf{diam}\left(C_j\right)$$

$$\leqslant \|q - \tilde{\mu}_j\| + \delta\left(\mathsf{diam}\,(C_i) + \mathsf{diam}\left(C_j\right)\right) - \frac{\alpha - 1}{\alpha}\|q - \mu_j\| \qquad \text{(Lemma 2.6)}$$

$$\leqslant \|q - \tilde{\mu}_j\| + \left(\delta - \frac{(\alpha - 1)}{4}\right)\left(\mathsf{diam}\,(C_i) + \mathsf{diam}\left(C_j\right)\right) \qquad \text{(Using (7))}$$

$$< \|q - \tilde{\mu}_j\|. \qquad\qquad \text{(Using (2))}$$

□

Proof of Theorem 1.6. Using Proposition 3.5 for a desired clustering $C_1, \ldots, C_k$, we get a set of points $\{\tilde{\mu}_1, \ldots, \tilde{\mu}_k\}$, such that

$$\|\tilde{\mu}_i - \mu_i\| \leqslant \delta\, \mathsf{diam}\,(C_i) \ \text{ for all } i \in [k].$$

Using the Lemma 2.8 we get that all the points are closer to their approximate means than other approximate means. This step is achieved by Algorithm 4. Using the Lemma 2.10 we get that even with respect to the approximate means, the outlier points are the farthest points of the data-set. Therefore we can remove the farthest $|Z|$ points, and get the desired clustering. This step is achieved by Algorithm 3. □

Proof of Theorem 1.7. The proof immediately follows from the Proposition 3.5. We think of β points as outliers for the Algorithm 4. With a constant probability, we get that one of the clusterings in the list produced by the algorithm corresponds to $(n - \beta)$ points being α-center proximal. □

3 Sampling

3.1 Balanced Cluster Assumption

The following proposition works for general $A_1, \ldots, A_k, Z$. For our case, A_i corresponds to cluster C_i, and Z corresponds to the set of outliers.

Proposition 3.1. *Fix $\delta, \beta_1 \in (0, 1)$. Let X be a set of points in $\mathbb{R}^d$, partitioned into $k + 1$ sets $A_1, \ldots, A_k, Z$ such that for each i, $|A_i| \geqslant \omega |X|/k$. Let μ_i denote the mean of A_i, i.e., $\mu_i \stackrel{\text{def}}{=} \left(\sum_{v \in A_i} v\right)/|A_i|$. There exists a randomized algorithm which outputs the set $\left\{(\tilde{\mu}_1^{(p)}, \ldots, \tilde{\mu}_k^{(p)}) : p \in [T]\right\}$ satisfying*

$$\mathbb{P}\left[\exists p \in [T] \text{ such that } \left\|\tilde{\mu}_i^{(p)} - \mu_i\right\| \leqslant \delta\, \mathsf{diam}\,(A_i) \text{ for each } i \in [k]\right] \geqslant 1 - \beta_1,$$

Algorithm 4: List of α-center proximal k-means with balanced clusters with outliers

Input: a set of points $X = \{x_1, x_2, \ldots, x_n\} \subseteq \mathbb{R}^d$, a positive integer $k \in \mathbb{Z}_{\geqslant 0}$, a number $\alpha > 1$, $\omega > 0$, and $z \geqslant 0$

Output: A list $\mathcal{L}$ of k-tuples, where each k-tuple k-mean points of a clustering.

1: $\delta \leftarrow (\alpha - 1)^2 / 8\alpha$
2: Initialize $(\tilde{\mu}_1^{(0)}, \tilde{\mu}_2^{(0)}, \ldots, \tilde{\mu}_k^{(0)}) \leftarrow (\bar{0}, \bar{0}, \ldots, \bar{0})$
3: $m \leftarrow (12/\omega) \, (128k^2\alpha^2 / (\alpha - 1)^4 \beta_1) \, \log(2k/\beta_1)$
4: $\mathcal{L} \leftarrow \emptyset$
5: Pick an i.i.d. and uniform sample Y of size m from X.
6: **if** z = 0 **then**
7:     **for** enumeration $p \in [T]$ over the set T of all k disjoint partitions $Y = Y_1^{(p)} \cup Y_2^{(p)} \ldots \cup Y_k^{(p)}$ **do**
8:         **for** $j = 1$ to k **do**
9:             $\tilde{\mu}_j^{(p)} \leftarrow \frac{1}{\left|Y_j^{(p)}\right|} \sum_{x_i \in Y_j^{(p)}} x_i$
10:         **end for**
11:         $\mathcal{L} = \mathcal{L} \cup \left\{ \left(\tilde{\mu}_1^{(p)}, \ldots, \tilde{\mu}_k^{(p)} \right) \right\}$
12:     **end for**
13: **end if**
14: **if** $z > 0$ **then**
15:     **for** enumeration $p \in [T]$ over the set T of all $k + 1$ disjoint partitions $Y = Y_1^{(p)} \cup Y_2^{(p)} \ldots \cup Y_{k+1}^{(p)}$ **do**
16:         **for** $j = 1$ to k **do**
17:             $\tilde{\mu}_j^{(p)} \leftarrow \frac{1}{\left|Y_j^{(p)}\right|} \sum_{x_i \in Y_j^{(p)}} x_i$
18:         **end for**
19:         $\mathcal{L} = \mathcal{L} \cup \left\{ \left(\tilde{\mu}_1^{(p)}, \ldots, \tilde{\mu}_k^{(p)} \right) \right\}$
20:     **end for**
21: **end if**

in $O(T)$ iterations where $T < (k+1)^{\frac{16k^2}{\delta^2 \beta_1} \log\left(\frac{2k}{\beta_1}\right)}$.

We begin by recalling the following lemma, which is implicit in Theorem 2 of [Bar15], which says that with constant probability, one can get close to the mean of a set of points with bounded diameter, by randomly sampling a constant number of points.

Lemma 3.2. *Fix a set of elements A and a set of parameters $\delta, \beta_3 \in (0,1)$, and let $l_0 = 1/\left(\delta^2 \beta_3\right)$. Let $l \in \mathbb{Z}$ be any number such that $l \geqslant l_0$, and let $y_1, \ldots y_l$ be l independent and uniformly random samples from A. Then,*

$$\mathbb{P}\left[\left\|\frac{1}{l}\sum_{i=1}^{l} y_i - \frac{1}{|A|}\sum_{a\in A} a\right\| \leqslant \delta \operatorname{diam}(A)\right] \geqslant 1 - \beta_3.$$

Claim 3.3.

$$\mathbb{E}\left\|\frac{1}{l}\sum_{i=1}^{l} y_i - \frac{1}{|A|}\sum_{a\in A} a\right\|^2 \leqslant \operatorname{diam}(A)^2/l$$

Proof. (We are reproducing the proof of [Bar15] for the sake of completeness.)

$$\mathbb{E}\left\|\frac{1}{l}\sum_{i=1}^{l} y_i - \frac{1}{|A|}\sum_{a\in A} a\right\|^2 = \frac{1}{l^2}\mathbb{E}\left\|\sum_{i=1}^{l}\left(\frac{1}{|A|}\sum_{a\in A} a - y_i\right)\right\|^2$$

$$= \frac{1}{l^2}\mathbb{E}\left\langle\sum_{i=1}^{l}\left(\frac{1}{|A|}\sum_{a\in A} a - y_i\right), \sum_{j=1}^{l}\left(\frac{1}{|A|}\sum_{a\in A} a - y_j\right)\right\rangle$$

$$= \frac{1}{l^2}\sum_{i,j=1}^{l}\mathbb{E}\left\langle\left(\frac{1}{|A|}\sum_{a\in A} a - y_i\right), \left(\frac{1}{|A|}\sum_{a\in A} a - y_j\right)\right\rangle$$

$$= \frac{1}{l^2}\sum_{i=1}^{l}\mathbb{E}\left\langle\left(\frac{1}{|A|}\sum_{a\in A} a - y_i\right), \left(\frac{1}{|A|}\sum_{a\in A} a - y_i\right)\right\rangle \qquad \left(\text{Since } y_i \text{ and } y_j \text{ are independent}\right)$$

$$= \frac{1}{l^2}\sum_{i=1}^{l}\left(\left\|\frac{1}{|A|}\sum_{a\in A} a\right\|^2 - 2\left\|\frac{1}{|A|}\sum_{a\in A} a\right\|^2 + \mathbb{E}\|y_i\|^2\right) \qquad \left(\mathbb{E}\, y_i = \frac{1}{|A|}\sum_{a\in A} a\right)$$

$$\leqslant \frac{1}{l^2} l \operatorname{diam}(A)^2 \qquad \left(\mathbb{E}\|y_i\|^2 - \|x\|^2 \leqslant \operatorname{diam}(A)^2\right)$$

$$= \frac{\operatorname{diam}(A)^2}{l}$$

$\square$

Proof of Lemma 3.2. Let the random variable $Z \overset{\text{def}}{=} \left\|\frac{1}{l}\sum_{i=1}^{l} y_i - \frac{1}{|A|}\sum_{a\in A} a\right\|^2$. We know from the Claim 3.3 that $\mathbb{E}(Z) \leqslant \frac{\operatorname{diam}(A)^2}{l}$. Therefore using Markov's inequality, we get

$$\mathbb{P}\left[Z \geqslant \frac{\operatorname{diam}(A)^2}{l\beta_3}\right] \leqslant \frac{\mathbb{E}\, Z}{\frac{\operatorname{diam}(A)^2}{l\beta_3}} \leqslant \frac{\frac{\operatorname{diam}(A)^2}{l}}{\frac{\operatorname{diam}(A)^2}{l\beta_3}} = \beta_3$$

$$\mathbb{P}\left[\left|\sqrt{Z}\right| \geqslant \frac{\operatorname{diam}(A)}{\sqrt{l}\sqrt{\beta_3}}\right] \leqslant \beta_3$$

Therefore choosing $\delta = \frac{1}{\sqrt{\beta_3}\sqrt{l}}$, we get the statement of Lemma 3.2. $\square$

The above lemma helps us determine the number of points to be taken from each of the clusters to get close to the mean. The next helps us determine the number of points to sample uniformly at random from the data-set, such that we get at least some fixed number of points from each balanced clusters (with a balance parameter ω), with constant probability.

Lemma 3.4. *Fix $l_0 \in \mathbb{Z}$ and $\beta_2 \in (0, 1/2]$. Let $X = \{x_1, \ldots, x_n\}$ be a set of n distinct items partitioned into $k + 1$ sets $A_1, \ldots, A_k, Z$ such that $|A_i| \geqslant \omega n / k$ for each $i \in [k]$. Let $m = (8/\omega)kl_0 \log(k/\beta_2)$, and let $Y = \{y_1, \ldots, y_m\}$ be a set of m independently and uniformly randomly chosen elements from X (with repetition). Then,*

$$\mathbb{P}\left[|Y \cap A_i| \geqslant l_0 \text{ for each } i \in [k]\right] \geqslant 1 - \beta_2 .$$

Proof. Let $p_i \overset{\text{def}}{=} |A_i| / n$ and let $Y_i \overset{\text{def}}{=} |Y \cap A_i|$. For our choice of parameters, we get that

$$\frac{p_i m}{2} \geqslant \frac{\omega m}{2k} \geqslant l_0 \qquad \text{and} \qquad \frac{p_i m}{8} \geqslant \log \frac{k}{\beta_2} .$$

Therefore, using the Chernoff bound on Y_i, we get that

$$\mathbb{P}[Y_i \geqslant l_0] \geqslant \mathbb{P}[Y_i \geqslant p_i m / 2] \geqslant 1 - e^{-p_i m / 8} \geqslant 1 - \frac{\beta_2}{k} .$$

Using a union bound over each $i \in [k]$, we get

$$\mathbb{P}[|Y \cap A_i| \geqslant l_0 \text{ for each } i \in [k]] \geqslant 1 - k\frac{\beta_2}{k} = 1 - \beta_2 .$$

$\square$

We are now ready to prove Proposition 3.1.

Proof of Proposition 3.1. Set $\beta_2 = \beta_1/2$, $\beta_3 = \beta_1/(2k)$, set l_0 from the guarantee in Lemma 3.2, and set m from the guarantee in Lemma 3.4. Let $Y = \{y_1, \ldots, y_m\}$ be a set of m independently and uniformly randomly chosen elements from X (with repetition) It is easy to verify that the points in $Y \cap A_i$ are a uniformly random sample from A_i.

$$\mathbb{P}\left[y_j = a \mid y_j \in A_i\right] = \frac{1}{|A_i|} \qquad \text{for any } i \in [k], a \in A_i \text{ and } j \in [m] \tag{8}$$

Let $\tilde{\mu}_i^{(*)} \overset{\text{def}}{=} \left(\sum_{j \in Y \cap A_i} y_j\right) / |Y \cap A_i|$. If we enumerate all the $k + 1$ partitions on Y, then one of the partitions will be $\{Y \cap A_1, \ldots, Y \cap A_k, Y \cap Z\}$. For a fixed index $i \in [k]$, using Lemma 3.4 and (8) with $A = A_i$ and $Y \cap A_i$ as the set of random samples from A_i, we get that

$$\mathbb{P}[|Y \cap A_i| \geqslant l_0 \text{ for each } i \in [k]] \geqslant 1 - \beta_2 = 1 - \frac{\beta_1}{2} .$$

In this case, using Lemma 3.2, we get that

$$\mathbb{P}\left[\left\|\tilde{\mu}_i^{(*)} - \mu_i\right\| \leqslant \delta \operatorname{diam}(A_i) \text{ for each } i \in [k]\right] \geqslant 1 - k\beta_3 = 1 - \frac{\beta_1}{2} .$$

Using a union bound over these two events, we get that

$$\mathbb{P}\left[\left\|\tilde{\mu}_i^{(*)} - \mu_i\right\| \leqslant \delta \operatorname{diam}(A_i) \text{ for each } i \in [k]\right] \geqslant 1 - \frac{\beta_1}{2} - \frac{\beta_1}{2} = 1 - \beta_1 .$$

The running time of this algorithm is dominated by the time required to enumerate all the $k + 1$ partitions of set of cardinality m, and computing the means of those partitions. $\square$

3.2 Balanced Mean Distance Assumption

Assume that the unknown α-center proximal k-means clustering of lowest cost is $\{C_1, \ldots, C_k\}$, with means $\{\mu_1, \ldots, \mu_k\}$ respectively. In this section we assume that the ratio of the maximum pairwise distance between the means to the minimum pairwise distance between the means is bounded by a factor γ^*. We assume that we are given an upper bound γ on γ^*. More formally, we are given a γ, such that

$$\gamma \geqslant \frac{\max_{i,j} \|\mu_i - \mu_j\|}{\min_{i \neq j} \|\mu_i - \mu_j\|}.$$

Proposition 3.5. *Fix $\delta \in (0, 1)$. Let X be a set of points in $\mathbb{R}^d$, partitioned into k sets $C_1, \ldots, C_k$, and let μ_i denote the mean of C_i, i.e., $\mu_i \overset{\text{def}}{=} \left(\sum_{v \in C_i} v \right) / |C_i|$, such that $\gamma = \max_{i,j} \|\mu_i - \mu_j\| \big/ \min_{i \neq j} \|\mu_i - \mu_j\|$. Let $\varepsilon = \left(\frac{2\delta k}{\gamma(1+k)} \right)^2$. The Algorithm 5 constructs $O\left(2^{\mathsf{poly}\left(\frac{k}{\varepsilon}\right)}\right)$ k-tuples. With constant probability, there exists at least one k-tuple $(\tilde{\mu}_1, \ldots, \tilde{\mu}_k)$ satisfying*

$$\|\tilde{\mu}_i - \mu_i\| \leqslant 2\delta r_{i,j} \text{ for all } j \in [k] \text{ and for each } i \in [k].$$

Moreover, the algorithm runs in time $O\left(2^{\mathsf{poly}\left(\frac{k}{\varepsilon}\right)} nd\right)$.

. The proof of the Proposition 3.5 is similar to the proof of Lemma 3.3 of [DX15], with minor modifications, keeping our application in mind.

Using Proposition 2.3 we define $\mathsf{rad}_{\min}(X) \overset{\text{def}}{=} \frac{\alpha}{\alpha^2 - 1} \min_{i \neq j} \|\mu_i - \mu_j\|$. The proposition implies that the radius of the largest $C_{i,j}$ ball (for some $i, j \in [k]$) is at most $\gamma \, \mathsf{rad}_{\min}(X)$, ie.,

$$\frac{r_{i,j}}{\gamma} \leqslant \mathsf{rad}_{\min}(X) \leqslant r_{i,j} \qquad \text{for } i, j \in [k]. \tag{9}$$

Proposition 3.6. *Let $R \overset{\text{def}}{=} \max_{i,j \in [n]} \|x_i - x_j\|$, and let $\mathsf{rad}_{\min}(X) \overset{\text{def}}{=} \frac{\alpha}{\alpha^2 - 1} \min_{i \neq j} \|\mu_i - \mu_j\|$. Then,*

$$\mathsf{rad}_{\min}(X) \in \left[\left(\frac{\alpha}{(\alpha + 1)^2} \right) \frac{R}{\gamma}, \left(\frac{\alpha}{\alpha^2 - 1} \right) \frac{R}{\gamma} \right]. \tag{10}$$

Proof. It is easy to see that

$$R \geqslant \max_{i,j} \|\mu_i - \mu_j\|, \qquad \text{for } i, j \in [k],$$

and from Proposition 2.3 that (points lie in two balls of radius $r_{i,j}$ which are at a distance $d_{i,j}$ apart)

$$R \leqslant \max_{i,j} \left(4r_{i,j} + d_{i,j} \right) = \left(\frac{\alpha + 1}{\alpha - 1} \right) \max_{i,j} \|\mu_i - \mu_j\|, \qquad \text{for } i, j \in [k].$$

Using the above two equations and the fact that $\gamma = \max_{i,j} \|\mu_i - \mu_j\| \big/ \min_{i \neq j} \|\mu_i - \mu_j\|$, we get that

$$\min_{i \neq j} \|\mu_i - \mu_j\| \in \left[\left(\frac{\alpha - 1}{\alpha + 1} \right) \frac{R}{\gamma}, \frac{R}{\gamma} \right], \qquad \text{for } i, j \in [k],$$

which implies that (using Proposition 2.3)

$$\mathsf{rad}_{\min}(X) \in \left[\left(\frac{\alpha}{(\alpha + 1)^2} \right) \frac{R}{\gamma}, \left(\frac{\alpha}{\alpha^2 - 1} \right) \frac{R}{\gamma} \right].$$

$\square$

The Algorithm 6 is almost same as the Algorithm Peeling-and-Enclosing-Tree of [DX15], with a minor variation in step 2(b).

We note a set of preliminary lemmas and definition which we will need for the proof of Proposition 3.5.

Algorithm 5: Algorithm Peeling-and-Enclosing

Input: $X = \{x_1, \ldots, x_n\}$ in $\mathbb{R}^d$, $k \geqslant 2$, α, γ.
Output: A list $\mathcal{L}$ containing k-tuples, where a k-tuple contains k mean points.

1. Set $\varepsilon = \left(\frac{(\alpha-1)^2 k}{4\alpha\gamma(1+k)} \right)^2$.

2. For $i = 0$ to $\log_{(1+\varepsilon)} \left(\frac{\alpha+1}{\alpha-1} \right)$ do

    (a) $\zeta = (1 + \varepsilon)^i \left(\frac{\alpha}{(\alpha+1)^2} \right) \frac{R}{\gamma}$.

    (b) Run **Algorithm Peeling-and-Enclosing-Tree**.

    (c) Let $\mathcal{T}_i$ be the output tree.

3. For each root-to-leaf path of every $\mathcal{T}_i$, build a k-tuple candidate using the k points associated with the path.

4. Append the k-tuple to the list $\mathcal{L}$.

Algorithm 6: Algorithm Peeling-and-Enclosing-Tree

Input: ζ and an instance of k-means X.
Output: A tree $\mathcal{T}$.

1. Initialize $\mathcal{T}$ as a single root node v associated with no point.

2. Recursively grow each node v in the following way

    (a) If the height of v is already k, then it is a leaf.

    (b) Otherwise, let j be the height of v. Build the radius candidate set $\mathcal{R} = \left\{ \frac{1+l\frac{\varepsilon}{2}}{2(1+\varepsilon)} j \sqrt{2} \sqrt{\varepsilon} \zeta \gamma \middle| 0 \leqslant l \leqslant 4 + \frac{2}{\varepsilon} \right\}$
    For each $r \in \mathcal{R}$, do

        i. Let $\left\{ p_{v_1}, \ldots, p_{v_j} \right\}$ be the j points associated with nodes on the root-to-v path.
        ii. For each p_{v_l}, $1 \leqslant l \leqslant j$, construct a ball $\mathcal{B}_{j+1,l}$ centered at p_{v_l} and with radius r.
        iii. Take a random sample from $\left(X \setminus \cup_{l=1}^{j} \mathcal{B}_{j+1,l} \right)$ of size $s = \frac{8k^3}{\varepsilon^9} \ln \frac{k^2}{\varepsilon^6}$. Compute the mean points of all subset of the sample, and denote them by $\Pi = \{\pi_1, \ldots, \pi_{2^s-1}\}$.
        iv. For each $\pi_i \in \Pi$, construct a simplex using $\left\{ p_{v_1}, \ldots, p_{v_j}, \pi_i \right\}$ as its vertices. Also construct another simplex using $\left\{ p_{v_1}, \ldots, p_{v_j} \right\}$ as its vertices. For each simplex, build a grid with size $O\left((32j/\varepsilon)^j \right)$ inside itself and each of its 2^j possible degenerated sub-simplices.
        v. In total, there are $2^{s+j} (32j/\varepsilon)^j$ grid points inside the 2^s simplices. For each grid point, add one child to v, and associate it with the grid point.

3. Output $\mathcal{T}$.

Definition 3.7 (Simplex). A k-simplex is a k-dimensional polytope which is the convex hull of its $k+1$ vertices. More formally, suppose the $k+1$ points $u_0, \ldots, u_k \in \mathbb{R}^k$ are affinely independent, Then, the simplex determined by them is the set of points

$$\mathcal{V} = \left\{ \theta_0 u_0 + \cdots + \theta_k u_k \ \middle| \ \sum_{i=0}^{k} \theta_i = 1 \text{ and } \theta_i \geqslant 0 \text{ for all } i \in [k] \right\} .$$

Lemma 3.8 (Lemma 1, [IKI94]). *Let S be a set of n points in $\mathbb{R}^d$, T be a randomly selected subset of size t from S, and $\mu(S), \mu(T)$ be the mean points of S and T respectively. With probability $1 - \eta$, $\|\mu(S) - \mu(T)\|^2 \leqslant \frac{1}{\eta t} \sigma^2$, where $\sigma^2 = \frac{1}{n} \sum_{s \in S} \|s - \mu(s)\|^2$ and $0 \leqslant \eta \leqslant 1$.*

Lemma 3.9 (Lemma 4, [DX14]). *Let Γ be a set of elements, and S be a subset of Γ with $\frac{|S|}{|\Gamma|} = \rho$ for some $\rho \in (0, 1)$. If we randomly select $\frac{t \ln \frac{t}{\eta}}{\ln(1+\rho)} = O\left(\frac{t}{\rho} \ln \frac{t}{\eta}\right)$ elements from Γ, then with probability at least $1 - \eta$, the sample contains t or more elements from S for $0 < \eta < 1$ and $t \in \mathbb{Z}^+$.*

Lemma 3.10 (Lemma 2.3 (Simplex Lemma II), [DX15]). *Let Q be a set of points in $\mathbb{R}^d$ with a partition of $Q = \cup_{l=1}^{j} Q_l$ and $Q_{l_1} \cap Q_{l_2} = \emptyset$ for any $l_1 \neq l_2$. Let o be the mean point of Q, and o_l be the mean point of Q_l for $1 \leqslant l \leqslant j$. Let $\sigma^2 = \frac{1}{|Q|} \sum_{q \in Q} \|q - o\|^2$. Let $\{o'_1, \ldots, o'_j\}$ be j points in $\mathbb{R}^d$ such that $\|o_l - o'_l\| \leqslant L$ for $1 \leqslant l \leqslant j, L > 0$, and $\mathcal{V}'$ be the simplex determined by $\{o'_1, \ldots, o'_j\}$. Then for any $0 < \varepsilon \leqslant 1$, it is possible to construct a grid of size $O\left((8j/\varepsilon)^j\right)$ inside $\mathcal{V}'$ such that at least one grid point τ satisfies the inequality $\|\tau - o\| \leqslant \sqrt{\varepsilon} \sigma + (1 + \varepsilon) L$.*

Lemma 3.11 (Lemma 2.2, [DX15]). *Let Q be a set of points in $\mathbb{R}^d$, and Q_1 be its subset containing $\rho |Q|$ points for some $0 < \rho \leqslant 1$. Let o and o_1 be the mean points of $\mathbb{Q}$ and Q_1, respectively. Then $\|o - o_1\| \leqslant \sqrt{\frac{1-\rho}{\rho}} \sigma$, where $\sigma^2 = \frac{1}{|Q|} \sum_{q \in Q} \|q - o\|^2$.*

Notations: Let $\mathcal{OPT} = \{C_1, \ldots, C_k\}$ be the k unknown optimal clusters for the lowest cost α-center proximal k-means objective, with means μ_j. W.l.o.g. we assume that $|C_1| \geqslant \ldots \geqslant |C_k|$. We define $\sigma_j^2 \overset{\text{def}}{=} \frac{1}{|C_j|} \sum_{p \in C_j} \|p - \mu_j\|^2$. Let $\lambda_j \overset{\text{def}}{=} |C_j|/n$.

The following lemma is similar to the Lemma 3.3 of [DX15] with minor modifications.

Lemma 3.12. *Among all the points generated by the Algorithm 6, with constant probability, there exists at least one tree, $\mathcal{T}_i$, which has a root-to-leaf path with each of its nodes v_j at level $j, (1 \leqslant j \leqslant k)$ associating with a point p_{v_j} and satisfying the inequality*

$$\left\| p_{v_j} - \mu_j \right\| \leqslant \varepsilon \gamma \, \mathsf{rad}_{\min}(X) + (1 + \varepsilon) j \sqrt{\varepsilon} \gamma \, \mathsf{rad}_{\min}(X) . \tag{11}$$

Algorithm and Proof Overview: We will give a high level idea of the algorithm and the proof. At each searching step, the algorithm performs a 'sphere peeling' and 'simplex enclosing' step, to generate k approximate mean points for the clusters. Initially the algorithm uses a random sampling technique to find an approximate mean p_{v_1} for C_1. This can be done as $\frac{|C_1|}{n} \geqslant 1/k$, and hence we can sample. For some $j \geqslant 1$, suppose that at the $(j+1)^{th}$ iteration, the algorithm already has approximate mean points mean points $p_{v_1}, \ldots, p_{v_j}$ for $C_1, \ldots, C_j$. It is not clear how to distinguish points which belong to $C_1, \ldots, C_j$ from those which belong to C_{j+1}. Also, the number of points in the cluster C_{j+1} could be small, it is tough to obtain a significant fraction of such points using random sampling. Therefore, the idea used is to seperate the points in C_{j+1} using j peeling spheres, $\mathcal{B}_{j+1,1}, \ldots, \mathcal{B}_{j+1,j}$, centered at the j approximate mean points respectively and with radius approximately being $\mathsf{rad}_{\min}(X)$. Note that $\mathcal{B}_{j+1,1}, \ldots, \mathcal{B}_{j+1,j}$ can have some points from C_{j+1}. Let P_{j+1} be the set of unknown points in $C_{j+1} \setminus \left(\cup_{l=1}^{j} \mathcal{B}_{j+1,l} \right)$. The algorithm considers two cases, a) $|P_{j+1}|$ is large and b) $|P_{j+1}|$ is small. For the case a) when $|P_{j+1}|$ is large, we can sample points from P_{j+1} using random sampling, and get an approximate mean π of P_{j+1}, and then construct a simplex determined by $\pi, p_{v_1}, \ldots, p_{v_j}$ to contain the $(j+1)^{th}$ mean point, using Lemma 3.10. This is because, $C_{j+1} \cap \mathcal{B}_{j+1,l}, l \in [j]$ can be seen as a partition of C_{j+1} whose approximate mean is p_{v_l}, thus the simplex lemma II applies. For case b) where $|P_{j+1}|$ is small, it directly constructs the simplex determined by $p_{v_1}, \ldots, p_{v_j}$, and searches for the approximate mean point of C_{j+1} in the grid. This follows

18

because it can be shown that $C_{j+1} \cap \mathcal{B}_{j+1,l}$, $l \in [j]$ can be seen as a partition of C_{j+1} whose approximate mean is p_{v_l}, and from the Lemma 3.11. The Lemma 3.11 roughly says that even if we remove a small number of points from a cluster, its new mean remains close to the original mean.

Proof of Lemma 3.12: Let $\mathcal{T}_i$ be the tree generated by the Algorithm 6 when

$$\zeta \in [\, \text{rad}_{\min}(X), (1+\varepsilon)\,\text{rad}_{\min}(X)\,] \, .$$

We will prove this lemma by induction.

Base Case: For $j = 1$, we have $\lambda_1 \geqslant \frac{1}{k}$. Therefore through random sampling (Lemma 3.9), we can find a point p_{v_1}, which is close to μ_1 (Lemma 3.8). We get that $\left\| p_{v_1} - \mu_1 \right\| \leqslant \varepsilon\sigma_1 \leqslant \varepsilon\,\text{rad}_{\min}(X) + (1+\varepsilon)\gamma\sqrt{\varepsilon}\,\text{rad}_{\min}(X)$. Hence the base case holds.

Induction Step: We assume that there is a path in $\mathcal{T}_i$ from root to the $(j-1)$-level, such that for each $1 \leqslant l \leqslant (j-1)$, the level-$l$ node v_l on the path associated with a point p_{v_l} satisfying the inequality

$$\left\| p_{v_l} - \mu_l \right\| \leqslant \varepsilon\gamma\,\text{rad}_{\min}(X) + (1+\varepsilon)l\sqrt{\varepsilon}\gamma\,\text{rad}_{\min}(X) \, .$$

Now we need to show this for the j-level, i.e., we need to show that there exists at least one child v_j of v_{j-1}, such that the associated point p_{v_j} satisfies the inequality

$$\left\| p_{v_j} - \mu_j \right\| \leqslant \varepsilon\gamma\,\text{rad}_{\min}(X) + (1+\varepsilon)j\sqrt{\varepsilon}\gamma\,\text{rad}_{\min}(X) \, .$$

First we make the following claim. The claim is a slight modification of Claim 2 of [DX15]. We will prove it in Section 3.2.1.

Claim 3.13. *In the set of radius candidates in the algorithm, there exists one value $r_j \in \mathcal{R}$, such that*

$$r_j \in \left[j\sqrt{\varepsilon}\gamma\,\text{rad}_{\min}(X), \left(1 + \frac{\varepsilon}{2}\right)j\sqrt{\varepsilon}\gamma\,\text{rad}_{\min}(X) \right] \, .$$

Now we construct $(j-1)$ peeling spheres $\left\{ \mathcal{B}_{j,1}, \ldots, \mathcal{B}_{j,j-1} \right\}$. For each $1 \leqslant l \leqslant j-1$, $\mathcal{B}_{j,l}$ is centered at p_{v_l} with radius r_j. Next we make the following claim. The proof claim is similar to the proof of Claim 3 of [DX15], adapted to our setting. We will prove it in Section 3.2.1.

Claim 3.14. *For each $1 \leqslant l \leqslant j-1$, $\left| C_l \setminus \left(\cup_{w=1}^{j-1} \mathcal{B}_{j,w} \right) \right| \leqslant \frac{4\lambda_j n}{\varepsilon}$.*

Claim 3.14 shows that $\left| C_l \setminus \left(\cup_{w=1}^{j-1} \mathcal{B}_{j,w} \right) \right|$ is bounded for $1 \leqslant l \leqslant j-1$, which helps us to find the approximate mean of C_j. C_j is divided into j subsets, $(C_j \cap \mathcal{B}_{j,1}), \ldots, (C_j \cap \mathcal{B}_{j,j-1})$, and $C_j \setminus \left(\cup_{w=1}^{j-1} \mathcal{B}_{j,w} \right)$. Let P_l denote $C_j \cap \mathcal{B}_{j,l}$ for $1 \leqslant l \leqslant j-1$, and P_j denote $C_j \setminus \left(\cup_{w=1}^{j-1} \mathcal{B}_{j,w} \right)$, and τ_l denote mean point of P_l for $1 \leqslant l \leqslant j$. We can assume that $\{P_l | 1 \leqslant l \leqslant j\}$ are pairwise disjoint. If not, then arbitrarily assign points to either of the peeling spheres which intersect in C_j.

We now have two cases: (a) $\left| P_j \right| \geqslant \varepsilon^3 \frac{\lambda_j}{j} n$, and (b) $\left| P_j \right| < \varepsilon^3 \frac{\lambda_j}{j} n$. We show that Algorithm 6 can obtain an approximate mean for C_j by using the Lemma 3.10, for both the cases.

Case (a): By Claim 3.14, and using that fact that $\lambda_l \leqslant \lambda_j$ for $l > j$, we know that

$$\frac{\left| P_j \right|}{\sum_{1 \leqslant l \leqslant k} \left| C_l \setminus \left(\cup_{w=1}^{j-1} \mathcal{B}_{j,w} \right) \right|} \geqslant \frac{\frac{\varepsilon^3}{j}\lambda_j}{\frac{4(j-1)\lambda_j}{\varepsilon} + \lambda_j + (k-j)\lambda_j},$$

$$\geqslant \frac{\varepsilon^4}{8kj},$$

$$\geqslant \frac{\varepsilon^4}{8k^2} \, .$$

This means that P_j is large enough, compared to the points outside the peeling spheres. Hence we can use random sampling technique to obtain an approximate mean point π for P_j in the following way. First we set $t = \frac{k}{\varepsilon^5}$, $\eta = \frac{\varepsilon}{k}$, and take sample of size $\frac{8k^3}{\varepsilon^9} \ln \frac{k^2}{\varepsilon^6}$. By Lemma 3.9 we know that with probability $1 - \frac{\varepsilon}{k}$, the sample contains $\frac{k}{\varepsilon^5}$ points from P_j. Let π be the mean of the $\frac{k}{\varepsilon^5}$ points sampled from P_j, and let a^2 be the variance of P_j. By Lemma 3.8 we know that with probability $1 - \frac{\varepsilon}{k}$, $\left\| \pi - \tau_j \right\|^2 \leqslant \varepsilon^4 a^2$. Also, since $\frac{|P_j|}{|C_j|} = \frac{|P_j|}{\lambda_j n} \geqslant \frac{\varepsilon^3}{j}$, we have $a^2 \leqslant \frac{|C_j|}{|P_j|} \sigma_j^2 \leqslant \frac{1}{\varepsilon^3} \sigma_j^2$. This upper bound on a^2 follows because $P_j \subset C_j$. We are summing the distance square over all the elements in C_j, and then dividing by $|P_j|$. Thus, $\left\| \pi - \tau_j \right\|^2 \leqslant \varepsilon j \sigma_j^2 \leqslant \varepsilon j \gamma^2 \, \mathsf{rad}_{\min}(X)^2$.

After obtaining the point π, we can use the Lemma 3.10 to find a point p_{v_j} satisfying the condition of $\left\| p_{v_j} - \mu_j \right\| \leqslant \varepsilon \gamma \, \mathsf{rad}_{\min}(X) + (1 + \varepsilon) j \sqrt{\varepsilon} \gamma \, \mathsf{rad}_{\min}(X)$. This is true because of the following. First we construct the simplex $\mathcal{V}'_{(a)}$ with vertices $\left\{ p_{v_1}, \ldots, p_{v_{j-1}}, \pi \right\}$, which is $o'_1, \ldots o'_{j-1}, o'_j$ in the lemma. Note that C_j which is Q in the lemma is partitioned by the peeling spheres into j disjoint subsets $P_1, \ldots, P_j$, which is $Q_1, \ldots, Q_j$ in the lemma. Each P_l ($1 \leqslant l \leqslant j-1$) locates inside $\mathcal{B}_{j,l}$, which implies that τ_l (mean of P_l), which is o_l in the lemma, is also inside $\mathcal{B}_{j,l}$. Further by Claim 3.13 we have for $1 \leqslant l \leqslant j - 1$:

$$ j \sqrt{\varepsilon} \gamma \, \mathsf{rad}_{\min}(X) \leqslant r_j \leqslant \left(1 + \frac{\varepsilon}{2} \right) j \sqrt{\varepsilon} \gamma \, \mathsf{rad}_{\min}(X) \,. \tag{12} $$

We also have from above that (recall that τ_l is the mean of P_l)

$$ \left\| \pi - \tau_j \right\| \leqslant \sqrt{\varepsilon j} \gamma \, \mathsf{rad}_{\min}(X) \,. \tag{13} $$

By (12) and (13) we know that if we set the value of L and ε (in Lemma 3.10) to be $L = \left(1 + \frac{\varepsilon}{2} \right) j \sqrt{\varepsilon} \gamma \, \mathsf{rad}_{\min}(X)$ and ε to be $\varepsilon_0 = \varepsilon^2/4$, by Lemma 3.10 we can construct a grid inside the simplex $\mathcal{V}'_{(a)}$ with size $O\left((8j/\varepsilon_0)^j \right)$ to ensure existence of one grid point τ satisfying that inequality

$$ \left\| \tau - \mu_j \right\| \leqslant \sqrt{\varepsilon_0} \sigma_j + (1 + \varepsilon_0) L \leqslant \varepsilon \gamma \, \mathsf{rad}_{\min}(X) + (1 + \varepsilon) j \sqrt{\varepsilon} \gamma \, \mathsf{rad}_{\min}(X) \,. $$

Hence we can use τ as p_{v_j}, and the induction step holds.

Case (b): We can use the Lemma 3.10 to find an approximate mean point. This is true because of the following. We construct a simplex $\mathcal{V}'_{(b)}$ with vertices $\left\{ p_{v_1}, \ldots, p_{v_{j-1}} \right\}$. Since $|P_j|$ is small, the mean points of $C_j \setminus P_j$ and C_j are very close to each other (Lemma 3.11). Thus we can ignore P_j and consider only $C_j \setminus P_j$. Here the value of ρ in Lemma 3.11 is $(1 - \varepsilon^3/j)$. Thus, the $(j - 1)$ dimensional simplex will approximate the mean of P_j well (by Lemma 3.10). Here the value of L and ε is same as in the case (a). Therefore, the induction step holds for this case as well.

Since Algorithm 6 executes every step in the above discussion, the induction step, as well as the lemma, is true.

$\square$

Proof of Proposition 3.5. From the Lemma 3.12 we know that with constant probability our algorithm can find a point $\tilde{\mu}_j$, $1 \leqslant j \leqslant k$ such that

$$ \left\| \tilde{\mu}_j - \mu_j \right\| \leqslant \varepsilon \gamma \, \mathsf{rad}_{\min}(X) + (1 + \varepsilon) j \sqrt{\varepsilon} \gamma \, \mathsf{rad}_{\min}(X) \,. $$

Therefore we now calculate the value of ε, the success probability and the running time.

ε **Value:** From (11) (which implies that $\left\| \tilde{\mu}_j - \mu_j \right\| \leqslant \varepsilon \gamma \, \mathsf{rad}_{\min}(X) + (1 + \varepsilon) k \sqrt{\varepsilon} \gamma \, \mathsf{rad}_{\min}(X)$) and (9) (which gives a bound on $\mathsf{rad}_{\min}(X)$). From proof of Claim 3.14 we know that $\varepsilon \leqslant \frac{1}{4k^2}$. We want to show that $\left\| \tilde{\mu}_j - \mu_j \right\| \leqslant 2\delta r_{i,j}$. Therefore, we get that for $i, j \in [k]$

$$ (\varepsilon + (1 + \varepsilon) k \sqrt{\varepsilon}) \gamma r_{i,j} \leqslant 2 \delta r_{i,j} \,, $$

$$ \sqrt{\varepsilon} (\sqrt{\varepsilon} + (1 + \varepsilon) k) \leqslant \frac{2\delta}{\gamma} \,, $$

$$ \varepsilon \leqslant \left(\frac{2\delta k}{\gamma(1+k)} \right)^2 \qquad\qquad \left(\text{Since } \varepsilon \leqslant \frac{1}{4k^2} \right). \tag{14} $$

Therefore setting $\varepsilon \leqslant \left(\frac{2\delta k}{\gamma(1+k)} \right)^2$ we get statement of the proposition.

Success Probability: From the above analysis, we know that only in the **case (a)** in the analysis of Lemma 3.12 needs sampling. We took a sample of size $s = \frac{8k^3}{\varepsilon^9} \ln \frac{k^2}{\varepsilon^6}$. With probability $1 - \frac{\varepsilon}{k}$, it contains $\frac{k}{\varepsilon^5}$ points from P_j. Meanwhile, with probability, $1 - \frac{\varepsilon}{k}$, $\|\pi - \tau_j\|^2 \leq \varepsilon^4 a^2$. Hence the success probability in the jth iteration is $\left(1 - \frac{\varepsilon}{k}\right)^2$. Therefore the success probability in k iterations is $\left(1 - \frac{\varepsilon}{k}\right)^{2k} \geq 1 - 2\varepsilon$.

Runtime Analysis: Each node in the returned tree by Algorithm 6 has $|\mathcal{R}| \, 2^{s+j} \left(\frac{32j}{\varepsilon^2}\right)^j$ children, where $|\mathcal{R}| = O(1/\varepsilon)$, and $s = \frac{8k^3}{\varepsilon^9} \ln \frac{k^2}{\varepsilon^6}$. Since the tree has height k, the number of candidate points for the means are $O\left(2^{\mathsf{poly}\left(\frac{k}{\varepsilon}\right)}\right)$. Since each node takes $O\left(|\mathcal{R}| \, 2^{s+j} \left(\frac{32j}{\varepsilon^2}\right)^j nd\right)$ time, the time complexity of the Algorithm 6 is $O\left(2^{\mathsf{poly}\left(\frac{k}{\varepsilon}\right)} nd\right)$. The Algorithm 6 is called by the Algorithm 5 $O\left(\log_{(1+\varepsilon)}\left(\frac{\alpha+1}{\alpha-1}\right)\right)$ times. Therefore, the total running time of the Algorithm 5 is $O\left(2^{\mathsf{poly}\left(\frac{k}{\varepsilon}\right)} nd\right)$. $\qquad\square$

3.2.1 Proof of Claims

Proof of Claim 3.13. We know that

$$2^{-1/2} \sqrt{\varepsilon}\gamma \, \mathsf{rad}_{\min}(X) \leq \sqrt{\varepsilon}\gamma \, \mathsf{rad}_{\min}(X) \leq 2^{1/2} \sqrt{\varepsilon}\gamma \, \mathsf{rad}_{\min}(X) \, .$$

Together with the fact that $\zeta \in [\, \mathsf{rad}_{\min}(X), (1+\varepsilon) \, \mathsf{rad}_{\min}(X)]$, we get that

$$\frac{\sqrt{2}}{2} \sqrt{\varepsilon} \frac{\zeta\gamma}{(1+\varepsilon)} \leq \sqrt{\varepsilon}\gamma \, \mathsf{rad}_{\min}(X) \leq \sqrt{2} \sqrt{\varepsilon} \, \zeta\gamma \, .$$

Let $\hat{r}_j = \sqrt{2} \sqrt{\varepsilon} \, \zeta\gamma$, we get that

$$\sqrt{\varepsilon}\gamma \, \mathsf{rad}_{\min}(X) \leq \hat{r}_j \leq 2(1+\varepsilon) \sqrt{\varepsilon}\gamma \, \mathsf{rad}_{\min}(X) \, .$$

Let $z = \frac{j \, \hat{r}_j}{j \, \sqrt{\varepsilon}\gamma \, \mathsf{rad}_{\min}(X)}$. Then we have $1 \leq z \leq 2(1+\varepsilon)$. We build a grid in the interval $\left[\frac{z}{2(1+\varepsilon)}, z\right]$ with grid length $\frac{\varepsilon}{4(1+\varepsilon)} z$, and obtain a number set $\mathcal{N} = \left\{ \frac{1+l\frac{\varepsilon}{2}}{2(1+\varepsilon)} z \,\middle|\, 0 \leq l \leq 4 + \frac{2}{\varepsilon} \right\}$. We prove that there must exist one number in $\mathcal{N}$ and is between 1 and $1 + \varepsilon/2$. First, we know that $\frac{z}{2(1+\varepsilon)} \leq 1 \leq z$. If $z \leq 1 + \varepsilon/2$, we find the desired number in $\mathcal{N}$. Otherwise, the whole interval $[1, 1 + \varepsilon/2]$ is inside $\left[\frac{z}{2(1+\varepsilon)}, z\right]$. Since each grid has length $\frac{\varepsilon}{4(1+\varepsilon)} z \leq \frac{\varepsilon}{4(1+\varepsilon)} 2(1+\varepsilon) = \varepsilon/2$, there must exist one grid point located inside $[1, 1 + \varepsilon/2]$. Thus the desired number exists in $\mathcal{N}$.

$\mathcal{R} = \left\{ \frac{1+l\frac{\varepsilon}{2}}{2(1+\varepsilon)} j \, \hat{r}_j \,\middle|\, 0 \leq l \leq 4 + \frac{2}{\varepsilon} \right\}$, and from the above analysis we know that there exists one value $r_j \in \mathcal{R}$ such that

$$j \sqrt{\varepsilon}\gamma \, \mathsf{rad}_{\min}(X) \leq r_j \leq \left(1 + \frac{\varepsilon}{2}\right) j \sqrt{\varepsilon}\gamma \, \mathsf{rad}_{\min}(X) \, .$$

Note that $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Proof of Claim 3.14. For each $1 \leq l \leq j - 1$, we have that $\left|C_l \setminus \left(\cup_{w=1}^{j-1} \mathcal{B}_{j,w}\right)\right| \leq \left|C_l \setminus \mathcal{B}_{j,l}\right|$. By Markov's inequality we have

$$\left|C_l \setminus \mathcal{B}_{j,l}\right| \leq \frac{\sigma_l^2}{\left(r_j - \|p_{v_l} - \mu_l\|\right)^2} \, |C_l| \, .$$

The above inequality is true because suppose we take a random variable Y to be $\|x - \mu\|$. We know that $\mathbb{E} \, Y = \sigma_l^2$. Therefore using Markov's inequality for some $r_0 > 0$ we get that

$$\mathbb{P}\left[Y \geq r_0\right] \leq \frac{\sigma_l^2}{r_0} \, .$$

21

Multiplying the above equation by $|C_l|$, we get that the expected number of points outside the radius r_0 from center is less than $\sigma_l^2 \mod C_l / r_0$. Note that $\sigma_l^2 \leqslant \gamma^2 \operatorname{rad}_{\min}(X)^2$. Together with $r_j \geqslant j\sqrt{\varepsilon}\gamma \operatorname{rad}_{\min}(X)$ and $\left\| p_{v_l} - \mu_l \right\| \leqslant \varepsilon\gamma \operatorname{rad}_{\min}(X) + (1+\varepsilon)l\sqrt{\varepsilon}\gamma \operatorname{rad}_{\min}(X)$, we get

$$r_j - \left\| p_{v_l} - m_l \right\| \geqslant j\sqrt{\varepsilon}\gamma \operatorname{rad}_{\min}(X) - \left(\varepsilon\gamma \operatorname{rad}_{\min}(X) + (j-1)(1+\varepsilon)\sqrt{\varepsilon}\gamma \operatorname{rad}_{\min}(X)\right),$$

$$= (1 - (j-1)\varepsilon - \sqrt{\varepsilon})\sqrt{\varepsilon}\gamma \operatorname{rad}_{\min}(X).$$

Thus we have

$$\left| C_l \setminus \mathcal{B}_{j,l} \right| \leqslant \frac{\sigma_l^2}{(1 - (j-1)\varepsilon - \sqrt{\varepsilon})^2 \varepsilon\gamma^2 \operatorname{rad}_{\min}(X)^2} |C_l|,$$

$$\leqslant \frac{|C_l|}{(1 - (j-1)\varepsilon - \sqrt{\varepsilon})^2 \varepsilon}, \qquad\qquad \left(\sigma_l^2 \leqslant \gamma^2 \operatorname{rad}_{\min}(X)^2\right),$$

$$\leqslant \frac{\lambda_j n}{(1 - j\sqrt{\varepsilon})^2 \varepsilon}.$$

Note that we can assume ε is small enough such that $\varepsilon \leqslant \frac{1}{4k^2}$, which implies $\frac{\lambda_j n}{(1 - j\sqrt{\varepsilon})^2 \varepsilon} \leqslant \frac{4\lambda_j n}{\varepsilon}$, since $k \geqslant j$. Otherwise, we can just replace ε by $\frac{\varepsilon}{4k^2}$ as part of input at the beginning of the algorithm. Thus we have that

$$\left| C_l \setminus \mathcal{B}_{j,l} \right| \leqslant \frac{4\lambda_j n}{\varepsilon}.$$

$\square$

4 Lower Bound

4.1 Hardness Result

Our hardness result immediately follows from Awasthi et al. [ACKS15]. We will show a reduction from Vertex-Cover problem to the α-center proximal k-means clustering with balanced clusters. The Vertex-Cover problem can be stated as follows: Given an undirected graph $G = (V, E)$, choose a subset S of vertices with minimum $|S|$, such that S is incident on every edge of the graph. Awasthi et al. [ACKS15] showed the following lemma:

Lemma 4.1 (Corollary 5.3, [ACKS15])**.** *Given any unweighted triangle-free graph G with bounded degrees, it is NP-hard to approximate Vertex-Cover within any factor smaller than* 1.36.

Theorem 4.2. *There exists constants $\alpha > 1$, $\omega > 0$, $\varepsilon > 0$, such that there is an efficient reduction from instances of Vertex-Cover on triangle-free graphs of bounded degree to those of α-center proximal instances of Euclidean k-means clustering, where the size of each cluster is at least $\omega n/k$, that satisfies the following properties:*

(i) *if the Vertex-Cover instance has value k, the optimal α-center proximal k-means clustering where the size of each cluster is at least $\omega n/k$, has cost at most $m - k$.*

(ii) *if the Vertex-Cover instance has value at least $k(1 + \varepsilon)$, then the optimal α-center proximal k-means clustering, where the size of each cluster is at least $\omega n/k$, has a cost at least $m - (1 - \Omega(\varepsilon))k$.*

Proof. The construction of the k-means instance is same as that of [ACKS15]. Let $G = (V, E)$ denote the graph in the Vertex Cover instance $\mathcal{I}$, with parameter k denoting the number of vertices we can select. We assume that the graph G is triangle free and with a maximum degree $\Delta = \Omega(1)$. Let n be the number of vertices in the graph and m be the number of edges. We construct the k-means instance $\mathcal{I}_{km}$ as follows: for each vertex $i \in [n]$, we have a unit vector $x_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ which has 1 in the ith coordinate and 0 elsewhere. For each edge $e \equiv (i, j)$, we have a vector $x_e = x_i + x_j$. Our data points are $\{x_e : e \in E\}$.

Completeness: suppose $\mathcal{I}$ is such that there exists a vertex cover $S^* = \{v_1, \ldots, v_k\}$ of k vertices. We will show that we can recover an α-center proximal k-means clustering (where the size of each cluster is at least $\omega n/k$), of low cost.

Let E_{v_l} denote the set of edges covered by v_l for $1 \leqslant l \leqslant k$. If an edge is covered by two vertices, we will assume that only one of them covers it (arbitrarily). As a result each E_{v_l} is pairwise disjoint and their union is E.

We now do the clustering as follows. Consider a cluster $\mathcal{F}_v \overset{\text{def}}{=} \{x_e : e \in E_v\}$, which consists of data-points associated with edges covered by a single vertex v. Let $m_{\mathcal{F}_v}$ denote the number of edges v cover in the vertex cover, and let $\mu_{\mathcal{F}_v}$ denote the mean of $\mathcal{F}_v$. The mean $\mu_{\mathcal{F}_v}$ has a 1 in one of the coordinates (corresponding to x_v), $\left(1/\left|m_{\mathcal{F}_v}\right|\right)$ in $m_{\mathcal{F}_v}$ coordinates (corresponding to the edges), and 0 in the remaining.

Claim 4.3. *There exists an α-center proximal k-means clustering (where the size of each cluster is at least $\omega n/k$) of $\mathcal{I}_{km}$ with cost at most $m - k$, where m is the number of edges in the graph $G = (V, E)$ associated with the vertex cover instance $\mathcal{I}$, and k is the size of the optimal vertex cover.*

Proof. The cost of the cluster $\mathcal{F}_v$ is $\sum_{x \in \mathcal{F}_v} \left\| x - \mu_{\mathcal{F}_v} \right\|^2$. We note that for any $x \neq x' \in \mathcal{F}_v$, $\left\| x - \mu_{\mathcal{F}_v} \right\|^2 = \left\| x' - \mu_{\mathcal{F}_v} \right\|^2$. Therefore we get that the cost of the cluster $\mathcal{F}_v$ is:

$$m_{\mathcal{F}_v} \left((m_{\mathcal{F}_v} - 1) \left(\frac{1}{m_{\mathcal{F}_v}} \right)^2 + \left(1 - \frac{1}{m_{\mathcal{F}_v}} \right)^2 \right) = m_{\mathcal{F}_v} - 1 .$$

Summing this over the k cluster gives us the cost $m - k$.

Next, we bound the value of alpha for which this cluster is alpha stable. We note that the points closest to some other cluster center is the edge which was covered by two vertices. Let x_e be covered by v_1 and v_2. W.l.o.g. we assume that $x_e \in \mathcal{F}_{v_1}$. Let the number of edges in the cluster $\mathcal{F}_{v_1}$ be m_{v_1} and $\mathcal{F}_{v_2}$ be m_{v_2}, and let their respective means be μ_{v_1} and μ_{v_2}. The distance of x_e to mean of $\mathcal{F}_{v_1}$ is

$$\left\| x_e - \mu_{v_1} \right\|^2 = (m_{v_1} - 1) \left(\frac{1}{m_{v_1}} \right)^2 + \left(1 - \frac{1}{m_{v_1}} \right)^2 = \frac{m_{v_1} - 1}{m_{v_1}} ,$$

and the distance of x_e to mean of $\mathcal{F}_{v_2}$ is

$$\left\| x_e - \mu_{v_2} \right\|^2 = (m_{v_2}) \left(\frac{1}{m_{v_2}} \right)^2 + 1 = \frac{1 + m_{v_2}}{m_{v_2}} .$$

Note that the maximum degree of the graph is Δ. Therefore the value of α is

$$\alpha = \sqrt{\frac{m_{v_1}(m_{v_2} + 1)}{m_{v_2}(m_{v_1} - 1)}} \geqslant \sqrt{\frac{\Delta + 1}{\Delta - 1}} .$$

In the case where clusters do not share an edge, say $x_e \in \mathcal{F}_{v_1}$ we get that that (as per previous calculation)

$$\left\| x_e - \mu_{v_1} \right\|^2 = \frac{m_{v_1} - 1}{m_{v_1}} ,$$

and the distance of x_e to μ_{v_2} is

$$\left\| x_e - \mu_{v_1} \right\|^2 = 3 + (m_{v_2}) \left(\frac{1}{m_{v_2}} \right)^2$$

Therefore, we get that the value of α in this case is

$$\alpha = \sqrt{\frac{3 + \frac{1}{m_{v_2}}}{1 - \frac{1}{m_{v_1}}}} \geqslant \sqrt{\frac{\Delta + 3}{\Delta - 1}} > \sqrt{\frac{\Delta + 1}{\Delta - 1}} .$$

Next we bound the value of ω. The size of a cluster is bounded by the degree of the graph, ie., Δ, and the minimum size of a vertex cover for a graph with bounded degree Δ is $|S^*| \geqslant n/\Delta$. Therefore we get that the value of ω is $\frac{\omega n}{k} \geqslant 1$, $\omega \geqslant \frac{k}{n} \geqslant \frac{1}{\Delta}$. $\qquad \square$

Soundness: Next we show that if there is an α-center proximal k-means clustering, where the size of each cluster is at least $\omega n / k$, which has a low k-means cost, then there is a very good vertex cover for the corresponding graph. The proof for the soundness follows directly from Theorem 4.7 of [ACKS15].

Lemma 4.4 (Theorem 4.7, [ACKS15])**.** *If the k-means instance $\mathcal{I}_{km}$ has a clustering $\Gamma = \{\mathcal{F}_1, \dots, \mathcal{F}_k\} \sum_{\mathcal{F} \in \Gamma} Cost(\mathcal{F}) \leqslant m - (1 - \xi)k$, then there exists a $(1 + O(\delta))k$-vertex cover of G in the instance $\mathcal{I}$.*

The proof for our case follows from the above lemma because the statement holds for any k-means clustering of low cost, and hence it also holds for α-center proximal k-means clustering, where the size of each cluster is at least $\omega n / k$.

Combining Claim 4.3 and Lemma 4.4 we get the proof of Theorem 4.2. □

Proof of Theorem 1.5. From Theorem 4.2 we get that for some constant $\alpha > 1$, $\omega > 0$, $\varepsilon > 0$ if the vertex cover has value $k \geqslant m / \Delta$, then the α-center proximal k-means clustering, where the size of each cluster is at least $\omega n / k$, has cost at most $m\left(1 - \frac{1}{\Delta}\right)$, and if the vertex cover is at least $k(1 + \varepsilon)$, then optimal α-center proximal k means cost is at least $m\left(1 - \frac{1 - \Omega(\varepsilon)}{\Delta}\right)$. The vertex cover hardness Lemma 4.1 says that it is NP-hard to distinguish if the resulting α-center proximal k-means clustering, where the size of each cluster is at least $\omega n / k$, has cost at most $m\left(1 - \frac{1}{\Delta}\right)$ or cost more than $m\left(1 - \frac{1 - \Omega(\varepsilon)}{\Delta}\right)$. Since Δ is a constant, this implies that it is NP-hard to approximate α-center proximal k-means problem to within some factor $(1 + \Omega(\varepsilon))$, thereby proving the Theorem 1.5 □

4.2 On the Size of Possible Clustering

Proof of Proposition 1.9. Our construction is similar to the instance constructed by [BJK18] in their Theorem 2. We first construct the set of points X for which we fix an integer m such that $\alpha = \sqrt{1 + \frac{2}{m-1}} \leqslant \alpha'$. From this we get that $m = \frac{2}{\alpha^2 - 1} + 1$. The points in X belongs to $\mathbb{R}^d$, where $d = km$. The set X will have $n = d$ points : the standard basis vectors of $\mathbb{R}^d$, denoted by $e_1, \dots, e_d$. Now, we define the set of α-center proximal clusterings $\mathbb{C}$. The set $\mathbb{C}$ will consist of clusterings $C = \{C_1, \dots, C_k\}$, for which each of the clusters has exactly m points. We will now show that such a clustering is α-center proximal. Consider any two clusters, say C_1 and C_2. The mean μ_1 of C_1 has the value $1/m$ in its m coordinates and 0 in other coordinates. The μ_2 of C_2 has the value $1/m$ in its m coordinates and 0 in other coordinates, and for a non-zero coordinate of μ_1, μ_2 has 0 in the respective coordinate, since the clusters are disjoint by definition. Therefore, consider a point $x_i \in C_1$, we get that

$$\|x_1 - \mu_1\|^2 = \left(1 - \frac{1}{m}\right)^2 + (m - 1)\left(\frac{1}{m}\right)^2 = 1 - \frac{1}{m}.$$

and

$$\|x_1 - \mu_2\|^2 = 1 + m\left(\frac{1}{m}\right)^2 = 1 + \frac{1}{m}.$$

Therefore the value of α is

$$\alpha = \sqrt{\frac{1 + \frac{1}{m}}{1 - \frac{1}{m}}} = \sqrt{\frac{m+1}{m-1}}.$$

The number of such possible clusterings are

$$|\mathbb{C}| = \frac{(km)!}{(m!)^k} \approx k^{(km)}.$$

Therefore we get that the total possible number of such clusterings are

$$k^{k\left(\frac{2}{\alpha^2 - 1}\right) + 1} = 2^{\tilde{\Omega}\left(\frac{k}{\alpha^2 - 1}\right)}.$$

□

Acknowledgements

References

[ABD09]   Margareta Ackerman and Shai Ben-David, *Clusterability: A theoretical study*, Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics, Proceedings of Machine Learning Research, vol. 5, PMLR, 16–18 Apr 2009, pp. 1–8. 4

[ABS10]   Marcel R. Ackermann, Johannes Blömer, and Christian Sohler, *Clustering for metric and nonmetric distance measures*, ACM Trans. Algorithms **6** (2010), no. 4, 59:1–59:26. 4

[ABS12]   Pranjal Awasthi, Avrim Blum, and Or Sheffet, *Center-based clustering under perturbation stability*, Inf. Process. Lett. **112** (2012), no. 1-2, 49–54. 3

[ACKS15]   Pranjal Awasthi, Moses Charikar, Ravishankar Krishnaswamy, and Ali Kemal Sinop, *The Hardness of Approximation of Euclidean k-Means*, 31st International Symposium on Computational Geometry (SoCG 2015), Leibniz International Proceedings in Informatics (LIPIcs), vol. 34, Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2015, pp. 754–767. 1, 22, 24

[ADHP09]   Daniel Aloise, Amit Deshpande, Pierre Hansen, and Preyas Popat, *Np-hardness of euclidean sum-of-squares clustering*, Machine Learning **75** (2009), no. 2, 245–248. 1

[AMM17]   Haris Angelidakis, Konstantin Makarychev, and Yury Makarychev, *Algorithms for stable and perturbation-resilient problems*, Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, ACM, 2017, pp. 438–451. 3, 4

[ANFSW17] S. Ahmadian, A. Norouzi-Fard, O. Svensson, and J. Ward, *Better guarantees for k-means and euclidean k-median by primal-dual algorithms*, 2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS), Oct 2017, pp. 61–72. 4

[AS12]   Pranjal Awasthi and Or Sheffet, *Improved spectral-norm bounds for clustering*, Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, Springer Berlin Heidelberg, 2012, pp. 37–49. 5

[AV07]   David Arthur and Sergei Vassilvitskii, *K-means++: The advantages of careful seeding*, Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '07, Society for Industrial and Applied Mathematics, 2007, pp. 1027–1035. 4

[Bar15]   Siddharth Barman, *Approximating nash equilibria and dense bipartite subgraphs via an approximate version of caratheodory's theorem*, Proceedings of the Forty-seventh Annual ACM Symposium on Theory of Computing, STOC '15, ACM, 2015, pp. 361–369. 14

[BBG13]   Maria-Florina Balcan, Avrim Blum, and Anupam Gupta, *Clustering under approximation stability*, J. ACM **60** (2013), no. 2, 8:1–8:34. 5

[Ben15]   Shai Ben-David, *Computational feasibility of clustering under clusterability assumptions*, CoRR **abs/1501.00437** (2015). 1

[BHW16]   Maria-Florina Balcan, Nika Haghtalab, and Colin White, *k-Center Clustering Under Perturbation Resilience*, 43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016), Leibniz International Proceedings in Informatics (LIPIcs), vol. 55, Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2016, pp. 68:1–68:14. 3

[BJK18]   Anup Bhattacharya, Ragesh Jaiswal, and Amit Kumar, *Faster algorithms for the constrained k-means problem*, Theory of Computing Systems **62** (2018), no. 1, 93–115. 4, 24

[BL12]   Yonatan Bilu and Nathan Linial, *Are stable instances easy?*, Combinatorics, Probability and Computing **21** (2012), no. 5, 643660. 3

[BL16]   Maria-Florina Balcan and Yingyu Liang, *Clustering under perturbation resilience*, SIAM J. Comput. **45** (2016), no. 1, 102–155. 3, 5

[BR14]   Shalev Ben-David and Lev Reyzin, *Data stability in clustering: A closer look*, Theor. Comput. Sci. **558** (2014), 51–61. 3

[CAKM16] V. Cohen-Addad, P. N. Klein, and C. Mathieu, *Local search yields approximation schemes for k-means and k-median in euclidean and minor-free metrics*, 2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS), Oct 2016, pp. 353–364. 4

[CG18]   Chandra Chekuri and Shalmoli Gupta, *Perturbation Resilient Clustering for k-Center and Related Problems via LP Relaxations*, Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2018) (Dagstuhl, Germany) (Eric Blais, Klaus Jansen, José D. P. Rolim, and David Steurer, eds.), Leibniz International Proceedings in Informatics (LIPIcs), vol. 116, Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2018, pp. 9:1–9:16. 4

[Che08]   Ke Chen, *A constant factor approximation algorithm for k-median clustering with outliers*, Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms (Philadelphia, PA, USA), SODA '08, Society for Industrial and Applied Mathematics, 2008, pp. 826–835. 4

[Che09]   Ke Chen, *On coresets for k-median and k-means clustering in metric and euclidean spaces and their applications*, SIAM J. Comput. **39** (2009), no. 3, 923–947. 1, 4

[CKMN01] Moses Charikar, Samir Khuller, David M. Mount, and Giri Narasimhan, *Algorithms for facility location problems with outliers*, Proceedings of the Twelfth Annual ACM-SIAM Symposium on Discrete Algorithms (Philadelphia, PA, USA), SODA '01, Society for Industrial and Applied Mathematics, 2001, pp. 642–651. 4

[DF09]   S. Dasgupta and Y. Freund, *Random projection trees for vector quantization*, IEEE Transactions on Information Theory **55** (2009), no. 7, 3229–3242. 1

[DLS18]   Amit Deshpande, Anand Louis, and Apoorv Vikram Singh, *(1 + ε)-approximation for min-max k-means clustering*, Manuscript (2018). 4

[DX14]   Hu Ding and Jinhui Xu, *Sub-linear time hybrid approximations for least trimmed squares estimator and related problems*, Proceedings of the Thirtieth Annual Symposium on Computational Geometry (New York, NY, USA), SOCG'14, ACM, 2014, pp. 110:110–110:119. 18

[DX15]   _____ , *A unified framework for clustering constrained data without locality property*, Proceedings of the Twenty-sixth Annual ACM-SIAM Symposium on Discrete Algorithms (Philadelphia, PA, USA), SODA '15, Society for Industrial and Applied Mathematics, 2015, pp. 1471–1490. 4, 16, 18, 19

[FKRS18]   Zachary Friggstad, Kamyar Khodamoradi, Mohsen Rezapour, and Mohammad R. Salavatipour, *Approximation schemes for clustering with outliers*, Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms (Philadelphia, PA, USA), SODA '18, Society for Industrial and Applied Mathematics, 2018, pp. 398–414. 4

[FKS18]   Zachary Friggstad, Kamyar Khodamoradi, and Mohammad R. Salavatipour, *Exact algorithms and lower bounds for stable instances of euclidean k-means*, Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, 2018, pp. 2958–2972. 3, 4

[FMS07]   Dan Feldman, Morteza Monemizadeh, and Christian Sohler, *A ptas for k-means clustering based on weak coresets*, Proceedings of the Twenty-third Annual Symposium on Computational Geometry, SCG '07, ACM, 2007, pp. 11–18. 4

[FRS16]   Z. Friggstad, M. Rezapour, and M. R. Salavatipour, *Local search yields a ptas for k-means in doubling metrics*, 2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS), Oct 2016, pp. 365–374. 4

[GKL⁺17]   Shalmoli Gupta, Ravi Kumar, Kefu Lu, Benjamin Moseley, and Sergei Vassilvitskii, *Local search methods for k-means with outliers*, Proc. VLDB Endow. **10** (2017), no. 7, 757–768. 4

[HPK05]   Sariel Har-Peled and Akash Kushal, *Smaller coresets for k-median and k-means clustering*, Proceedings of the Twenty-first Annual Symposium on Computational Geometry, SCG '05, ACM, 2005, pp. 126–134. 4

[HPM04]   Sariel Har-Peled and Soham Mazumdar, *On coresets for k-means and k-median clustering*, Proceedings of the Thirty-sixth Annual ACM Symposium on Theory of Computing, STOC '04, ACM, 2004, pp. 291–300. 4

[IKI94]   Mary Inaba, Naoki Katoh, and Hiroshi Imai, *Applications of weighted voronoi diagrams and randomization to variance-based k-clustering: (extended abstract)*, Proceedings of the Tenth Annual Symposium on Computational Geometry, SCG '94, ACM, 1994, pp. 332–339. 4, 18

[JKS14]   Ragesh Jaiswal, Amit Kumar, and Sandeep Sen, *A simple d^2-sampling based ptas for k-means and other clustering problems*, Algorithmica **70** (2014), no. 1, 22–46. 4

[JV01]   Kamal Jain and Vijay V. Vazirani, *Approximation algorithms for metric facility location and k-median problems using the primal-dual schema and lagrangian relaxation*, J. ACM **48** (2001), no. 2, 274–296. 1

[KK10]   Amit Kumar and Ravindran Kannan, *Clustering with spectral norm and the k-means algorithm*, 51th Annual IEEE Symposium on Foundations of Computer Science (FOCS) 2010, October 23-26, 2010, 2010, pp. 299–308. 5

[KLS18]   Ravishankar Krishnaswamy, Shi Li, and Sai Sandeep, *Constant approximation for k-median and k-means with outliers via iterative rounding*, Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing (New York, NY, USA), STOC 2018, ACM, 2018, pp. 646–659. 4

[KMN⁺04]   Tapas Kanungo, David M. Mount, Nathan S. Netanyahu, Christine D. Piatko, Ruth Silverman, and Angela Y. Wu, *A local search approximation algorithm for k-means clustering*, Computational Geometry **28** (2004), no. 2, 89 – 112, Special Issue on the 18th Annual Symposium on Computational Geometry - SoCG2002. 4

[KSS04]   Amit Kumar, Yogish Sabharwal, and Sandeep Sen, *A simple linear time $(1+\varepsilon)$-approximation algorithm for k-means clustering in any dimensions*, 45th Symposium on Foundations of Computer Science (FOCS 2004), 17-19 October 2004, Rome, Italy, Proceedings, 2004, pp. 454–462. 1, 4

[MMV14]   Konstantin Makarychev, Yury Makarychev, and Aravindan Vijayaraghavan, *Bilulinial stable instances of max cut and minimum multiway cut*, Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, 2014, pp. 890–906. 3

[MNV12]   Meena Mahajan, Prajakta Nimbhorkar, and Kasturi Varadarajan, *The planar k-means problem is np-hard*, Theoretical Computer Science **442** (2012), 13 – 21, Special Issue on the Workshop on Algorithms and Computation (WALCOM 2009). 1

[TV10]   Matus Telgarsky and Andrea Vattani, *Hartigan's method: k-means clustering without voronoi*, Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, Proceedings of Machine Learning Research, vol. 9, PMLR, 13–15 May 2010, pp. 820–827. 3

[Vat11]   Andrea Vattani, *K-means requires exponentially many iterations even in the plane*, Discrete Comput. Geom. **45** (2011), no. 4, 596–616. 1

[VDW17]   Aravindan Vijayaraghavan, Abhratanu Dutta, and Alex Wang, *Clustering stable instances of euclidean k-means.*, Advances in Neural Information Processing Systems 30, Curran Associates, Inc., 2017, pp. 6503–6512. 5, 7

[VKKR03]   W. Fernandez de la Vega, Marek Karpinski, Claire Kenyon, and Yuval Rabani, *Approximation schemes for clustering problems*, Proceedings of the Thirty-fifth Annual ACM Symposium on Theory of Computing, STOC '03, ACM, 2003, pp. 50–58. 4