Hardness of Approximation for Euclidean k-Median

Anup Bhattacharya
School of Computer Sciences, National Institute of Science Education and Research (NISER), Khurda, India

Dishant Goyal
Indian Institute of Technology Delhi, India

Ragesh Jaiswal
Indian Institute of Technology Delhi, India

Abstract

The Euclidean k-median problem is defined in the following manner: given a set $\mathcal{X}$ of n points in d-dimensional Euclidean space $\mathbb{R}^d$, and an integer k, find a set $C \subset \mathbb{R}^d$ of k points (called centers) such that the cost function $\Phi(C, \mathcal{X}) \equiv \sum_{x \in \mathcal{X}} \min_{c \in C} \|x - c\|_2$ is minimized. The Euclidean k-means problem is defined similarly by replacing the distance with squared Euclidean distance in the cost function. Various hardness of approximation results are known for the Euclidean k-means problem [7, 29, 17]. However, no hardness of approximation result was known for the Euclidean k-median problem. In this work, assuming the *unique games conjecture* (UGC), we provide the hardness of approximation result for the Euclidean k-median problem in $O(\log k)$ dimensional space. This solves an open question posed explicitly in the work of Awasthi et al. [7].

Furthermore, we study the hardness of approximation for the Euclidean k-means/k-median problems in the bi-criteria setting where an algorithm is allowed to choose more than k centers. That is, bi-criteria approximation algorithms are allowed to output βk centers (for constant $\beta > 1$) and the approximation ratio is computed with respect to the optimal k-means/k-median cost. We show the hardness of bi-criteria approximation result for the Euclidean k-median problem for any $\beta < 1.015$, assuming UGC. We also show a similar hardness of bi-criteria approximation result for the Euclidean k-means problem with a stronger bound of $\beta < 1.28$, again assuming UGC.

2012 ACM Subject Classification Theory of computation → Facility location and clustering; Theory of computation → Approximation algorithms analysis

Keywords and phrases Hardness of approximation, bicriteria approximation, approximation algorithms, k-median, k-means

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2021.4

Category APPROX

Related Version *Full Version*: `https://arxiv.org/pdf/2011.04221.pdf` [9]

Acknowledgements Dishant Goyal would like to thank TCS Research Scholar Program.

1 Introduction

We start by giving the definition of the Euclidean k-median problem.

▶ **Definition 1** (k-median). *Given a set $\mathcal{X}$ of n points in d-dimensional Euclidean space $\mathbb{R}^d$, and a positive integer k, find a set of centers $C \subset \mathbb{R}^d$ of size k such that the cost function $\Phi(C, \mathcal{X}) \equiv \sum_{x \in \mathcal{X}} \min_{c \in C} \|x - c\|$ is minimized.*

The Euclidean k-means problem is defined similarly by replacing the distance with squared Euclidean distance in the cost function (i.e., replacing $\|x-c\|$ with $\|x-c\|^2$). These problems are also studied in the discrete setting where the centers are restricted to be chosen from a specific set $L \subset \mathbb{R}^d$, also given as input. This is known as the *discrete* version whereas the former version (with $L = \mathbb{R}^d$) is known as the *continuous* version. In the approximation

setting, the continuous version is not harder than its discrete counterpart since it is known(e.g., [22, 36]) that an α-approximation for the discrete problem gives an $\alpha + \varepsilon$ approximation for the continuous version, for arbitrary small constant $\varepsilon > 0$, in polynomial time. In this work, we only study the hardness of approximation for continuous version of the problem. The hardness of approximation for the discrete version thus follows from the hardness of approximation of continuous version. In the rest of the paper, we use k-means/median to implicitly mean *continuous Euclidean* k-means/median unless specified otherwise [1].

The relevance of the k-means and k-median problems in various computational domains such as resource allocation, big data analysis, pattern mining, and data compression is well known. A significant amount of work has been done to understand the computational aspects of the k-means/median problems. The k-means problem is known to be NP-hard even for fixed k or d [4, 20, 33, 40]. Similar NP hardness result is also known for the k-median problem [37]. Even the 1-median problem, popularly known as the *Fermat-Weber* problem [21], is a hard problem and designing efficient algorithms for this problem is a separate line of research in itself – see for e.g. [27, 42, 12, 10, 15]. These hardness barriers motivate approximation algorithms for these problems and a lot of progress have been made in this area. For example, there are various polynomial time approximation schemes (PTASs) known for k-means and k-median when k is fixed (or constant) [36, 28, 22, 14, 25]. Similarly, various PTASs are known for fixed d [19, 23, 16]. A number of constant factor approximation algorithms are also known for k-means and k-median when k and d are considered as part of the input. For the k-means problem, constant approximation algorithms have been given [26, 2], the best being a 6.357 approximation algorithm by Ahmadian *et al.* [2]. On the negative side, there exists a constant $\varepsilon > 0$ such that there does not exist an efficient $(1 + \varepsilon)$-approximation algorithm for the k-means problem, assuming $\mathsf{P} \neq \mathsf{NP}$ [7, 29, 17]. The best-known hardness of approximation result for the k-means problem is 1.07 due to Cohen-Addad and Karthik [17].

The constant factor approximation algorithms for the k-median problem are also known [13, 5, 32, 11, 2]. The best known approximation guarantee for k-median is 2.633 due to Ahmadian et al. [2]. On the hardness side, it was known that for general metric spaces, the discrete k-median problem is hard to approximate within a factor of $1 + 2/e$ [24]. However, unlike the Euclidean k-means problem, no hardness of approximation result was known for the Euclidean k-median problem. Resolving the hardness of approximation for the Euclidean k-median problem was left as an open problem in the work of Awasthi et al. [7]. They asked whether their techniques for proving the inpproximability results for Euclidean k-means can be used to prove the hardness of approximation result for the Euclidean k-median problem. From their paper,

> "*It would also be interesting to study whether our techniques give hardness of approximation results for the Euclidean k-median problem.*"

In this work, assuming UGC, we solve this open problem by obtaining the hardness of approximation result for the Euclidean k-median problem. Following is one of the main results of this work.

▶ **Theorem 2** (Main Theorem)**.** *There exists a constant $\varepsilon > 0$ such that the Euclidean k-median problem in $O(\log k)$ dimensional space cannot be approximated to a factor better than $(1 + \varepsilon)$, assuming the Unique Games Conjecture.*

[1] In some literature, the Euclidean space implicitly means the dimension is bounded, but in our case the dimension d can be arbitrarily large

Having established the hardness of approximation results for k-means and k-median, the next natural step in the discussion is to allow more flexibility to the algorithm. One possible relaxation is to allow an approximation algorithm to choose more than k centers, say, βk centers (for some constant $\beta > 1$) and produce a solution that is close to the optimal solution with respect to k centers. This is known as bi-criteria approximation and the following definition formalizes this notion.

▶ **Definition 3** ((α, β)-approximation algorithm). *An algorithm $\mathcal{A}$ is called an (α, β) approximation algorithm for the Euclidean k-means/k-median problem if given any instance $\mathcal{I} = (\mathcal{X}, k)$ with $\mathcal{X} \subset \mathbb{R}^d$, $\mathcal{A}$ outputs a center set $F \subset \mathbb{R}^d$ of size βk that has the cost at most α times the optimal cost with k centers. That is,*

$$\sum_{x \in \mathcal{X}} \min_{f \in F}\{D(x, f)\} \leq \alpha \cdot \min_{\substack{C \subseteq \mathbb{R}^d \\ |C|=k}} \left\{ \sum_{x \in \mathcal{X}} \min_{c \in C}\{D(x, c)\} \right\}$$

For the Euclidean k-means problem, $D(p, q) \equiv \|p - q\|^2$ and for the k-median problem $D(p, q) \equiv \|p - q\|$.

One expects that as β grows, there would exist efficient (α, β)-approximation algorithms with smaller values of α. This is indeed observed in the work of Makarychev et al. [35]. For example, their algorithm gives a $(9 + \varepsilon)$ approximation for $\beta = 1$; 2.59 approximation for $\beta = 2$; 1.4 approximation for $\beta = 3$. In other words, the approximation factor of their algorithm decreases as the value of β increases. Furthermore, their algorithm gives a $(1 + \varepsilon)$-approximation guarantee with $O(k \log(1/\varepsilon))$ centers. Bandyapadhyay and Varadarajan [8] gave a $(1 + \varepsilon)$ approximation algorithm that outputs $(1 + \varepsilon)k$ centers in constant dimension. There are various other bi-criteria approximation algorithms that use distance-based sampling techniques and achieve better approximation guarantees than their non bi-criteria counterparts [3, 1, 41]. Unfortunately in these bi-criteria algorithms, at least one of α, β is large. Ideally, we would like to obtain a PTAS with a small violation of the number of output centers. More specifically, we would like to address the following question:

> *Does the Euclidean k-means or Euclidean k-median problem admit an efficient $(1 + \varepsilon, 1 + \varepsilon)$-approximation algorithm?*

Note that such type of bi-criteria approximation algorithms that outputs $(1+\varepsilon)k$ centers have been extremely useful in obtaining a constant approximation for the *capacitated* k-median problem [30, 31] for which no true constant approximation is known yet.[2] Therefore, the above question is worth exploring. Note that here we are specifically aiming for a PTAS since the k-means and k-median problems already admit a constant factor approximation algorithm. In this work, we give a negative answer to the above question by showing that there exists a constant $\varepsilon > 0$ such that an efficient $(1 + \varepsilon, 1 + \varepsilon)$-approximation algorithm for the k-means and k-median problems does not exist assuming the Unique Games Conjecture. The following two theorems state this result more formally.

▶ **Theorem 4** (k-median). *For any constant $1 < \beta < 1.015$, there exists a constant $\varepsilon > 0$ such that there is no $(1 + \varepsilon, \beta)$-approximation algorithm for the Euclidean k-median problem in $O(\log k)$ dimensional space assuming the Unique Games Conjecture.*

[2] In the capacitated k-median/k-means problem there is an additional constraint on each center that it cannot serve more than a specified number of clients (or points).

▶ **Theorem 5** (k-means). *For any constant $1 < \beta < 1.28$, there exists a constant $\varepsilon > 0$ such that there is no $(1+\varepsilon, \beta)$-approximation algorithm for the Euclidean k-means problem in $O(\log k)$ dimensional space assuming the Unique Games Conjecture. Moreover, the same result holds for any $1 < \beta < 1.1$ under the assumption that $\mathsf{P} \neq \mathsf{NP}$.*

For simplicity, we present the proof of our results in $O(n)$ dimensional space. However, the results easily extend to $O(\log k)$ dimensional space using dimensionality reduction techniques of Makarychev *et al.* [34].

Important note: We would like to note that assuming $\mathsf{P} \neq \mathsf{NP}$, a similar hardness of approximation result for the Euclidean k-median problem using different techniques has been obtained independently by *Vincent Cohen-Addad, Karthik C. S., and Euiwoong Lee*. We came to know about their results through personal communication with the authors. Since their manuscript has not been published online yet, we are not able to add a citation to their work.

In the next subsection, we discuss the known results on hardness of approximation of the k-means and k-median problems in more detail.

1.1 Related Works

Guha and Khuller proved a $(1+\frac{2}{e})$ hardness of approximation result for the discrete k-median problem for the general metric spaces [24]. The first hardness of approximation result for the Euclidean k-means problem was given by Awasthi et al. [7]. They obtained their result using a reduction from Vertex Cover on triangle-free graphs of bounded degree Δ to the Euclidean k-means instances. Their reduction yields a $(1+\frac{\varepsilon}{\Delta})$ hardness factor for the k-means problem for some constant $\varepsilon > 0$. Lee et al. [29] showed the hardness of approximation of Vertex Cover on triangle-free graphs of bounded degree four. Using $\Delta = 4$, they obtained a 1.0013 hardness of approximation for the Euclidean k-means problem. Subsequently, Cohen-Addad and Karthik [17] improved the hardness of approximation to 1.07 using a modified reduction from the *vertex coverage problem* instead of a reduction from the vertex cover problem. Moreover, they also gave several improved hardness results for the discrete k-means/k-median problems in general and ℓ_p metric spaces. In their more recent work, they also improved the hardness of approximation results for the continuous k-means/k-median problem in general metric spaces [18].

Unlike the Euclidean k-means problem, no hardness of approximation result was known for the Euclidean k-median problem. In this work, we give hardness of approximation result for the Euclidean k-median problem assuming the Unique Game Conjecture. As mentioned earlier, in an unpublished work communicated to us through personal communication, *Vincent Cohen-Addad, Karthik C. S., and Euiwoong Lee* have independently obtained hardness of approximation result for the Euclidean k-median problem using different set of techniques and under the assumption that $\mathsf{P} \neq \mathsf{NP}$. They also gave bi-criteria hardness of approximation results in ℓ_∞-metric for the k-means and k-median problems [18]. We would like to point out that in the bi-criteria setting, our result is the first hardness of approximation result for the Euclidean k-means/k-median problem to the best of our knowledge.

1.2 Technical Overview and Contributions

Awasthi et al. [7] proved the first hardness of approximation result for the Euclidean k-means problem. Given any instance $\mathcal{I} = (\mathcal{X}, k)$ for the Euclidean k-means problem, they showed that there exists an $\epsilon > 0$ such that obtaining $(1+\epsilon)$-approximation for Euclidean k-means

is NP-hard. In this work we build on their techniques to prove the inapproximability result for the Euclidean k-median problem. First, we describe the reduction employed by Awasthi et al. for the Euclidean k-means problem and some related results.

> *Construction of k-means instance:* Let (G, k) be a hard Vertex Cover instance where the graph G has bounded degree Δ. Let n and m denote respectively the number of vertices and the number of edges in the graph. A k-means instance $\mathcal{I} \coloneqq (\mathcal{X}, k)$ with $\mathcal{X} \subset \mathbb{R}^n$ is constructed as follows. For every vertex $i \in V$, we have an n-dimensional vector $x_i \in \{0,1\}^n$, which has a 1 at i^{th} coordinate and 0 everywhere else. For each edge $e = (i,j) \in E$, a point $x_e \coloneqq x_i + x_j$ is defined in $\{0,1\}^n$. The set $\mathcal{X} \coloneqq \{x_e \mid e \in E\}$ with m points in $\mathbb{R}^n$ and parameter k define the k-means instance.

Awasthi et al. proved the following theorem based on the above construction [7].

▶ **Theorem 6** (Theorem 4.1 [7]). *There is an efficient reduction from vertex cover on bounded degree triangle-free graphs to the Euclidean k-means problem that satisfies the following properties:*

1. *If vertex cover of the instance is k, then there is a k-means clustering of cost at most $(m-k)$.*
2. *If vertex cover of the instance is at least $(1+\varepsilon)k$, then the cost of optimal k-means clustering is at least $(m-k+\delta k)$.*

Here, ε is some fixed constant > 0 and $\delta = \Omega(\varepsilon)$.

Awasthi et al. [7] used the following hardness result for the vertex cover problem on bounded degree triangle-free graphs.

▶ **Theorem 7** (Corollary 5.3 [7]). *Given any unweighted bounded degree triangle-free graph G, it is NP-hard to approximate Vertex Cover within any factor smaller than 1.36.*

Theorem 6 and Theorem 7 together imply that the Euclidean k-means problem is APX-hard. A formal statement for the same is given as follows (see Section 4 of [7] for the proof of this result).

▶ **Corollary 8.** *There exists a constant $\varepsilon' > 0$ such that it is NP-hard to approximate the Euclidean k-means problem to any factor better than $(1+\varepsilon')$.*

We would like to obtain a similar gap-preserving reduction for the Euclidean k-median problem. The first obstacle one encounters in this direction is that unlike the 1-mean problem, there does not exist a closed form expression for the 1-median problem, and hence we don't have an exact expression for the optimal 1-median cost. We overcome this barrier by obtaining good upper and lower bounds on the optimal 1-median cost and showing that these bounds suffice for our purpose. More concretely, to upper bound the optimal 1-median cost, we use the centroid as the 1-median and compute the 1-median cost with respect to the centroid. To obtain a lower bound on the 1-median cost of a cluster, we use a decomposition technique to break a cluster into smaller sub-clusters[3] for which we can compute exact or good approximate lower bounds on the 1-median cost. Here we use a simple observation that the optimal 1-median cost of a cluster is at least the sum of the optimal 1-median costs of the sub-clusters. For any sub-cluster that corresponds to a star graph, one can compute the

[3] Since a set of edges in a graph form a cluster of points in the reduction, we use the terms sub-graphs and sub-clusters interchangeably.

exact 1-median cost using our reduction. In order to bound the 1-median cost for sub-clusters that correspond to non-star graphs, we use the following observation crucially: the optimal 1-median cost is preserved under any transformation that preserves the pairwise distances. For non-star graphs, we first employ such a transformation that preserves the 1-median cost and then compute this cost exactly in the projected space. Note that this technique does not give exact 1-median cost for any arbitrary non-star graph, but works only for some *special families* of non-star graphs. The main idea of the decomposition technique is to ensure that only these kinds of non-star graphs are created in the decomposition process. The upper and lower bounds on the 1-median cost, as constructed in the above manner, are used in the completeness and soundness steps of the proof of the reduction, respectively.

The analysis for the completeness part of the reduction is relatively straightforward. If the vertex cover of a graph is k, then the edges of the graph can be divided into k star sub-graphs, each of which results in a star cluster in the k-median instance. The cost for this clustering with k star clusters can be found using the reduction easily.

In the proof for the soundness part of our reduction, we prove the contrapositive statement that assumes the k-median clustering cost to be bounded and proves that the vertex cover of the graph is not too large. Our analysis crucially depends on the relation between the vertex cover of a subgraph and the 1-median cost for that subgraph. More specifically, we need to answer the following question. Given a graph with r edges having vertex cover z, how does the optimal 1-median cost for that graph behave with respect to z. For example, for star graphs, $z = 1$ and the optimal 1-median cost of a star graph on r edges is exactly $\sqrt{r(r-1)}$. For any non-star graph with r edges, we first show that the optimal 1-median cost of the non-star graph is at least the optimal 1-median cost of a star graph with r edges. For any non-star graph F with r edges, we denote by $\delta(F)$ the *extra cost* of F, defined as the difference of the optimal 1-median cost of F and the optimal 1-median cost of a star graph with r edges. If we can figure out non-trivial lower bounds for $\delta(F)$ for different non-star graphs F, then we would be done. But, figuring out these non-trivial lower bounds that work for any non-star graph is quite a daunting prospect. The way we overcome this in our work is as follows. We characterize the non-star graphs as having maximum matching of size two or more than two, and for each, we relate the *extra cost* of 1-median clustering of that graph with the vertex cover of that graph. We show that the extra cost of a non-star sub-graph is proportional to the number of vertex-disjoint edges in the sub-graph. And since we assume the k-median cost to be bounded, the number of vertex disjoint edges is also bounded, giving a small vertex cover.

We need one more idea to finish the proof for the soundness part of the reduction. We call a cluster "singleton" if there is only one point in the cluster. Note that any such cluster would cost zero in a k-median clustering. If there are a large number of singleton clusters, say $t < k$, then they pay zero to the cost of the solution, even though those edges have vertex cover t. We prove a key lemma showing that for any *hard* instance of the vertex cover, the vertex cover of the sub-graph spanned by t singleton edges is at most $\frac{2t}{3}$. We combine these ideas to prove that if k-median clustering cost is bounded, the vertex cover of the graph cannot be too large.

We also prove the hardness of bi-criteria approximation results for Euclidean k-means and k-median problems. The hardness of bi-criteria approximation for Euclidean k-median is obtained by extending the proof for the hardness of approximation for the Euclidean k-median problem. We use the same reduction from the vertex cover problem and show that the soundness guarantees hold even if one is allowed to use βk centers, for some $\beta > 1$. We also show that similar techniques give the hardness of bi-criteria approximation results for the Euclidean k-means problem.

2 Useful Facts and Inequalities

In this section, we discuss some basic facts and inequalities that we will frequently use in our proofs. First, we note that the Fermat-Weber problem is not difficult for all 1-median instances. We can efficiently obtain 1-median for some special instances. For example, for a set of equidistant points, the 1-median is simply the centroid of the point set. We give a proof of this statement in the next section. Most importantly, we use the following fact and lemma to compute the 1-median cost.

▶ **Fact 1** ([38]). *For a set of non-collinear points the optimal* 1*-median is unique.*

▶ **Lemma 9.** *Let* $A = \{a_1, \ldots, a_n\}$ *and* $B = \{b_1, \ldots, b_n\}$ *be any two sets of* n *points in* $\mathbb{R}^d$*. If the pairwise distances between points within* A *is the same as pairwise distance between points within* B*. That is, for all* $i, j \in \{1, \ldots, n\}$*,* $\|a_i - a_j\| = \|b_i - b_j\|$*. Then the optimal* 1*-median cost of* A *is the same as the optimal* 1*-median cost of* B*.*

The proof of Lemma 9 is deferred to Appendix A. We use the above lemma, in vector spaces where it is tricky to compute the optimal 1-median exactly. In such cases, we transform the space to a different vector space, where computing the 1-median is relatively simpler. More specifically, we employ a rigid transformation since it preserves pairwise distances. Next, we give a simple lemma, that is used to prove various bounds related to the quantity $\sqrt{m(m-1)}$.

▶ **Lemma 10.** *Let* m *and* t *be any positive real numbers greater than one. If* $m \geq t$*, the following bound holds:*

$$m - (t - \sqrt{t(t-1)}) \leq \sqrt{m(m-1)} \leq m - 1/2.$$

Proof. The upper bound follows from the sequence of inequalities: $\sqrt{m(m-1)} < \sqrt{m^2 - m + 1/4} = \sqrt{(m-1/2)^2} = m - 1/2$. The lower bound follows from the following sequence of inequalities:

$$\sqrt{m(m-1)} = m + m \cdot \left(\sqrt{\frac{m-1}{m}} - 1 \right) \geq m + t \cdot \left(\sqrt{\frac{t-1}{t}} - 1 \right) = m - (t - \sqrt{t(t-1)}).$$

The second inequality holds because $\frac{a+1}{b+1} \geq \frac{a}{b}$ for $b \geq a$. This completes the proof of the lemma. ◀

2.1 Preliminaries

Recall that a point in $\mathcal{X}$ corresponds to an edge of the graph. Therefore, a sub-graph S of G corresponds to a subset of points $\mathcal{X}(S) \coloneqq \{x_e \mid e \in E(S)\}$ of $\mathcal{X}$. We define the 1-median cost of $\mathcal{X}(S)$ with respect to a center $c \in \mathbb{R}^n$ as $\Phi(c, S) \equiv \sum_{x \in \mathcal{X}(S)} \|x - c\|$. Furthermore, we define the **optimal** 1-median cost of $\mathcal{X}(S)$ as $\Phi^*(S)$. That is, $\Phi^*(S) \equiv \min_{c \in \mathbb{R}^n} \Phi(c, S)$. We often use these statements interchangeably, "optimal 1-median cost of a graph S" to mean "optimal 1-median cost of the cluster $\mathcal{X}(S)$".

3 Inapproximability of Euclidean k-Median

In this section, we show the inapproximability result of the Euclidean k-median problem. We obtain this result by showing a gap preserving reduction from Vertex Cover on bounded degree triangle-free graphs to the Euclidean k-median. For Vertex Cover on bounded degree triangle-free graphs, the inapproximability result is stated in Corollary 13. The corollary simply follows from the following two results of Austrin et al. [6] and Awasthi et al. [7].

▶ **Theorem 11** (Austrin et al. [6]). *Given any unweighted bounded degree graph $G = (V, E)$ of maximum degree Δ, Vertex Cover can not be approximated within any factor smaller than $2 - \varepsilon$, for $\varepsilon = (2 + o_\Delta(1)) \cdot \frac{\log\log\Delta}{\log\Delta}$ assuming the Unique Games Conjecture.*

In the above theorem, ε can be set to arbitrarily small value by taking sufficiently large value of Δ.

▶ **Theorem 12** (Awasthi et al. [7]). *There is a $(1 + \varepsilon)$-approximation-preserving reduction from Vertex Cover on bounded degree graphs to Vertex Cover on triangle-free graphs of bounded degree.*

▶ **Corollary 13.** *Given any unweighted triangle-free graph G of bounded degree, Vertex Cover can not be approximated within a factor smaller than $2 - \varepsilon$, for any constant $\varepsilon > 0$, assuming the Unique Games Conjecture.*

Earlier, in Section 1.2, we described the reduction used by Awasthi et al. [7] to construct instances for Euclidean k-means from a Vertex Cover instance. We use the same construction for the Euclidean k-median instances. Let $G = (V, E)$ denote a triangle-free graph of bounded degree Δ. Let $\mathcal{I} = (\mathcal{X}, k)$ denote the Euclidean k-median instance constructed from G. We establish the following theorem based on this construction.

▶ **Theorem 14.** *There is an efficient reduction from Vertex Cover on bounded degree triangle-free graphs with m edges to the Euclidean k-median problem that satisfies the following properties:*

1. *If the graph has a vertex cover of size k, then the k-median instance has a solution of cost at most $m - k/2$.*
2. *If the graph has no vertex cover of size at most $(2 - \varepsilon) \cdot k$, then the cost of any k-median solution on the instance is at least $m - k/2 + \delta k$.*

Here, ε is some fixed constant, $\delta = \Omega(\varepsilon)$, and $k \geq$ the size of maximum matching of the graph.

The graphs with a vertex cover of size at most k are said to be "Yes" instances and the graphs with no vertex cover of size at most $(2 - \varepsilon)k$ are said to be "No" instances. Now, the above theorem gives the following inapproximability result for the Euclidean k-median problem.

▶ **Corollary 15.** *There exists a constant $\varepsilon' > 0$ such that the Euclidean k-median problem can not be approximated to a factor better than $(1 + \varepsilon')$, assuming the Unique Games Conjecture.*

Proof. Since the hard Vertex Cover instances have bounded degree Δ, the maximum matching of such graphs is at least $\lceil \frac{m}{2\Delta} \rceil$. First, let us prove this statement. Suppose M be a matching, that is initially empty, i.e., $M = \emptyset$. We construct M in an iterative manner. First, we pick an arbitrary edge from the graph and add it to M. Then, we remove this edge and all the edges incident on it. We repeat this process for the remaining graph until the graph becomes empty. In each iteration, we remove at most 2Δ edges. Therefore, the matching size of the graph is at least $\lceil \frac{m}{2\Delta} \rceil$.

Now, suppose $k < \frac{m}{2\Delta}$. Then, the graph does not have a vertex cover of size k since matching size is at least $\lceil \frac{m}{2\Delta} \rceil$. Therefore, such graph instances can be classified as "No" instances in polynomial time. So, they are not the hard Vertex Cover instances. Therefore, we can assume $k \geq \frac{m}{2\Delta}$ for all the hard Vertex Cover instances. In that case, the second property of Theorem 14, implies that the cost of k-median instance is $(m - \frac{k}{2}) + \delta k \geq (1 + \frac{\delta}{2\Delta}) \cdot (m - \frac{k}{2})$. Thus, the k-median problem can not be approximated within any factor smaller than $1 + \frac{\delta}{2\Delta} = 1 + \Omega(\varepsilon)$. ◀

3.1 Completeness

Let $W = \{v_1, \ldots, v_k\}$ be a vertex cover of G. Let S_i denote the set of edges covered by v_i. If an edge is covered by two vertices v_i and v_j, then we arbitrarily keep the edge either in S_i or S_j. Let m_i denote the number of edges in S_i. We define $\{\mathcal{X}(S_1), \ldots, \mathcal{X}(S_k)\}$ as a clustering of the point set $\mathcal{X}$. Now, we show that the cost of this clustering is at most $m - k/2$. Note that each S_i forms a star graph centered at v_i. Moreover, the point set $\mathcal{X}(S_i)$ forms a regular simplex of side length $\sqrt{2}$. We compute the optimal cost of $\mathcal{X}(S_i)$ using the following lemma.

▶ **Lemma 16.** *For a regular simplex on r vertices and side length s, the optimal 1-median is the centroid of the simplex. Moreover, the optimal 1-median cost is $s \cdot \sqrt{\frac{r(r-1)}{2}}$.*

Proof. The statement is easy to see for $r = 1$. For $r = 2$, there are two points s distance apart. Therefore, the optimal center lies on the line segment joining the two points and the optimal 1-median cost is trivially s. So, for the rest of the proof, we assume that $r > 2$. Suppose $A = \{a_1, a_2, \ldots, a_r\}$ denote the vertex set of a regular simplex. Let s be the side length of the simplex. Using Lemma 9, we can represent each point a_i in an r-dimensional space as follows; we use the same notation to denote the points after such transformations.

$$a_1 \coloneqq \left(\frac{s}{\sqrt{2}}, 0, ..., 0\right), \quad a_2 \coloneqq \left(0, \frac{s}{\sqrt{2}}, ..., 0\right), \quad \ldots, \quad a_r \coloneqq \left(0, 0, ..., \frac{s}{\sqrt{2}}\right)$$

Note that the distance between any a_i and a_j is s, which is the side length of the simplex. Let $c^* = (c_1, \ldots, c_r)$ be an optimal 1-median of point set A. Then, the 1-median cost is the following:

$$\Phi(c^*, A) = \sum_{i=1}^{r} \|a_i - c^*\| = \sum_{i=1}^{r} \left(\sum_{j=1}^{r} c_j^2 - c_i^2 + \left(\frac{s}{\sqrt{2}} - c_i\right)^2 \right)^{1/2}$$

Suppose $c_i \neq c_j$ for any $i \neq j$. Then, we can swap c_i and c_j to create a different median, while keeping the 1-median cost the same. It contradicts the fact that there is only one optimal 1-median, by Fact 1. Therefore, we can assume $c^* = (c, c, \ldots, c)$. Now, the optimal 1-median cost is:

$$\Phi^*(A) = \Phi(c^*, A) \coloneqq r \cdot \sqrt{\left(c - \frac{s}{\sqrt{2}}\right)^2 + (r-1) \cdot c^2}$$

The function $\Phi(c^*, A)$ is strictly convex and attains minimum at $c = \frac{s}{m \cdot \sqrt{2}}$, which is the centroid of A. The optimal 1-median clustering cost is $\Phi(c^*, A) = s \cdot \sqrt{\frac{r(r-1)}{2}}$. This completes the proof of the lemma. ◀

The following corollary establishes the cost of a star graph S_i.

▶ **Corollary 17.** *Any star graph S_i with r edges has the optimal 1-median cost of $\sqrt{r(r-1)}$*

Using this corollary, we bound the optimal k-median cost of $\mathcal{X}$ as follows. Let $OPT(\mathcal{X}, k)$ denote the optimal k-median cost of $\mathcal{X}$. The following sequence of inequalities proves the first property of Theorem 14.

$$OPT(\mathcal{X}, k) \leq \sum_{i=1}^{k} \Phi^*(S_i) \overset{(Corollary\ 17)}{=} \sum_{i=1}^{k} \sqrt{m_i(m_i - 1)} \overset{(Lemma\ 10)}{\leq} \sum_{i=1}^{k} \left(m_i - \frac{1}{2}\right) = m - \frac{k}{2}.$$

3.2 Soundness

Now, we prove the second property of Theorem 14. For this, we prove the equivalent contrapositive statement: If the optimal k-median clustering of $\mathcal{X}$ has cost at most $\left(m - \frac{k}{2} + \delta k\right)$, for some constant $\delta > 0$, then G has a vertex cover of size at most $(2-\varepsilon)k$, for some constant $\varepsilon > 0$. Let $\mathcal{C}$ denote an optimal k-median clustering of $\mathcal{X}$. We classify its optimal clusters into two categories: (1) *star* and (2) *non-star*. Let $F_1, F_2, \ldots, F_t$ denote the non-star clusters, and $S_1, \ldots, S_{k-t}$ denote the star clusters. For any star cluster, the vertex cover size is exactly one. Moreover, using Corollary 17, the optimal 1-median cost of any star cluster with r edges is $\sqrt{r(r-1)}$. On the other hand, it may be tricky to exactly compute the vertex cover or the optimal cost of any non-star cluster. Suppose the optimal 1-median cost of a non-star cluster F on r edges is given as $\sqrt{r(r-1)} + \delta(F)$, where $\delta(F)$ denotes the *extra-cost* due to a non-star cluster F. Using this, we define $\delta(F)$ as the following:

$$\delta(F) \equiv \Phi^*(F) - \sqrt{|F|(|F|-1)}$$

The following lemmas bound the vertex cover of F in terms of $\delta(F)$.

▶ **Lemma 18.** *Any non-star cluster F with a maximum matching of size two has a vertex cover of size at most* $1.62 + \left(\sqrt{2}+1\right)\delta(F)$.

▶ **Lemma 19.** *Any non-star cluster F with a maximum matching of size at least three has a vertex cover of size at most* $1.8 + \left(\sqrt{2}+1\right)\delta(F)$.

These lemmas are the key to proving the main result. We discussed the main proof ideas earlier in Section 1.2; however, due to page-limit, the complete proof is deferred to the full version of the paper [9]. Now, let us see how these lemmas give a vertex cover of size at most $(2-\varepsilon)k$. Let us classify the star clusters into the following two sub-categories:

(a) Clusters composed of exactly one edge. Let these clusters be: $P_1, P_2, \ldots, P_{t_1}$.

(b) Clusters composed of at least two edges. Let these clusters be: $S_1, S_2, \ldots, S_{t_2}$.

Similarly, we classify the non-star clusters into the following two sub-categories:

(i) Clusters with a maximum matching of size two. Let these clusters be: $W_1, W_2, \ldots, W_{t_3}$

(ii) Clusters with a maximum matching of size at least three. Let these clusters be: $Y_1, Y_2, \ldots, Y_{t_4}$

Note that $t_1 + t_2 + t_3 + t_4$ equals k. Now, consider the following strategy of computing the vertex cover of G. Suppose, we compute the vertex cover for every cluster separately. Let C_i be any cluster, and $|VC(C_i)|$ denote the vertex cover size of C_i. Then, the vertex cover of G can be simply bounded in the following manner:

$$|VC(G)| \leq \sum_{i=1}^{t_1} |VC(P_i)| + \sum_{i=1}^{t_2} |VC(S_i)| + \sum_{i=1}^{t_3} |VC(W_i)| + \sum_{i=1}^{t_4} |VC(Y_i)|$$

However, we can obtain a vertex cover of smaller size using a slightly different strategy. In this strategy, we first compute a minimum vertex cover of all the clusters except single edge clusters $P_1, P_2, \ldots, P_{t_1}$. Suppose that vertex cover is VC'. Then we compute a vertex cover for $P_1, P_2, \ldots, P_{t_1}$. Now, let us see why this strategy gives a vertex cover of smaller size than before. Note that some vertices in VC' may also cover the edges in $P_1, \ldots, P_{t_1}$. Suppose there are t_1' clusters in $P_1, \ldots, P_{t_1}$ that remain uncovered by VC'. Without loss of generality, assume these clusters to be $P_1, \ldots, P_{t_1'}$. Now, the vertex cover of G is bounded in

the following manner:

$$\begin{aligned}|VC(G)| &\leq |VC\left(\cup_{i=1}^{t_1'} P_i\right)| + |VC'| \\ &= |VC\left(\cup_{i=1}^{t_1'} P_i\right)| + |VC\left((\cup_{j=1}^{t_2} S_j) \cup (\cup_{k=1}^{t_3} W_k) \cup (\cup_{l=1}^{t_4} Y_l)\right)| \\ &\leq |VC\left(\cup_{i=1}^{t_1'} P_i\right)| + \sum_{i=1}^{t_2} |VC(S_i)| + \sum_{i=1}^{t_3} |VC(W_i)| + \sum_{i=1}^{t_4} |VC(Y_i)|\end{aligned}$$

Now, we will try to bound the size of the vertex cover of $P_1 \cup ... \cup P_{t_1'}$. Note that we can cover all these single-edge clusters with t_1' vertices by choosing one vertex per cluster. However, it may be possible to obtain a vertex cover of smaller size if we collectively consider all these clusters. Suppose E_P denote the set of all edges in $P_1, \ldots, P_{t_1'}$ and V_P denote the vertex set spanned by them. We define a graph $G_P = (V_P, E_P)$. Further, suppose that M_P is a maximal matching of G_P. Then, it is easy to see that if $|M_P| \leq t_1'/3 + 4\delta k$ for some $\delta > 0$, we can simply pick both end-points of every edge in M_P, and it would give a vertex cover of G_P of size at most $2t_1'/3 + 8\delta k$. On the other hand, if $|M_P| > t_1'/3 + 4\delta k$, we show that the graph G admits a vertex cover of size at most $(2k - 2\delta k)$. This fact is mentioned in the following lemma. Due to page limit, the proof is deferred to the full version of the paper [9].

▶ **Lemma 20.** *Let $\delta > 0$ be any constant and G_P be as defined above. If G_P does not have a vertex cover of size $\leq (\frac{2t_1'}{3} + 8\delta k)$, then G has a vertex cover of size at most $(2k - 2\delta k)$.*

Based on the above lemma, we will assume that all single edge clusters can be covered with $(\frac{2t_1'}{3} + 8\delta k) \leq (\frac{2t_1}{3} + 8\delta k)$ vertices; otherwise the graph has a vertex cover of size at most $(2k - 2\delta k)$ and the soundness proof would be complete. Now, we bound the vertex cover of the entire graph in the following manner.

$$\begin{aligned}|VC(G)| &\leq |VC\left(\cup_{i=1}^{t_1'} P_i\right)| + |VC'| \\ &= |VC\left(\cup_{i=1}^{t_1'} P_i\right)| + |VC\left((\cup_{j=1}^{t_2} S_j) \cup (\cup_{k=1}^{t_3} W_k) \cup (\cup_{l=1}^{t_4} Y_l)\right)| \\ &\leq \sum_{i=1}^{t_1'} |VC(P_i)| + \sum_{i=1}^{t_2} |VC(S_i)| + \sum_{i=1}^{t_3} |VC(W_i)| + \sum_{i=1}^{t_4} |VC(Y_i)| \\ &\leq \left(\frac{2t_1}{3} + 8\delta k\right) + t_2 + \sum_{i=1}^{t_3} \left(\left(\sqrt{2}+1\right)\delta(W_i) + 1.62\right) + \sum_{i=1}^{t_4} \left(\left(\sqrt{2}+1\right)\delta(Y_i) + 1.8\right), \\ &\qquad\qquad \text{(using Lemmas 18, 19, and 20)} \\ &= (0.67)t_1 + 8\delta k + t_2 + (1.62)t_3 + (1.8)t_4 + \left(\sqrt{2}+1\right)\left(\sum_{i=1}^{t_3}\delta(W_i) + \sum_{i=1}^{t_4}\delta(Y_i)\right)\end{aligned}$$

Since the optimal cost $OPT(\mathcal{X}, k) = \sum_{j=1}^{k}\sqrt{m_j(m_j-1)} + \sum_{i=1}^{t_3}\delta(W_i) + \sum_{i=1}^{t_4}\delta(Y_i) \leq m - k/2 + \delta k$, we get $\sum_{i=1}^{t_3}\delta(W_i) + \sum_{i=1}^{t_4}\delta(Y_i) \leq m - k/2 + \delta k - \sum_{j=1}^{k}\sqrt{m_j(m_j-1)}$. We substitute this value in the previous equation, and get the following inequality:

$$|VC(G)| \leq (0.67)t_1 + 8\delta k + t_2 + (1.62)t_3 + (1.8)t_4 + \left(\sqrt{2}+1\right)\cdot\left(m - k/2 - \sum_{j=1}^{k}\sqrt{m_j(m_j-1)} + \delta k\right)$$

Using Lemma 10, we obtain the following inequalities:

1. For any cluster P_j with $|P_j| = 1$, we have $\sqrt{|P_j|\,(|P_j|-1)} \geq |P_j| - 1$
2. For any cluster S_j with $|S_j| \geq 2$, we have $\sqrt{|S_j|\,(|S_j|-1)} \geq |S_j| - (2-\sqrt{2})$
3. For any cluster W_j with $|W_j| \geq 2$, we have $\sqrt{|W_j|\,(|W_j|-1)} \geq |W_j| - (2-\sqrt{2})$
4. For any cluster Y_j with $|Y_j| \geq 3$, we have $\sqrt{|Y_j|\,(|Y_j|-1)} \geq |Y_j| - (3-\sqrt{6})$

We substitute these values in the previous equation, and get the following inequality:

$$|VC(G)| \leq (0.67)t_1 + 8\delta k + t_2 + (1.62)t_3 + (1.8)t_4 + \left(\sqrt{2}+1\right)\cdot\Bigg(m - k/2 - \sum_{j=1}^{t_1}(|P_j|-1) - \sum_{j=1}^{t_2}\left(|S_j| - (2-\sqrt{2})\right) - \sum_{j=1}^{t_3}\left(|W_j| - (2-\sqrt{2})\right) - \sum_{j=1}^{t_4}\left(|Y_j| - (3-\sqrt{6})\right) + \delta k\Bigg)$$

Since the number of edges $m = \sum_{j=1}^{t_1}|P_j| + \sum_{j=1}^{t_2}|S_j| + \sum_{j=1}^{t_3}|W_j| + \sum_{j=1}^{t_4}|Y_j|$, we get the following inequality:

$$|VC(G)| \leq (0.67)t_1 + 8\delta k + t_2 + (1.62)t_3 + (1.8)t_4 + \left(\sqrt{2}+1\right)\cdot\Big(-k/2 + t_1 + t_2\cdot\left(2-\sqrt{2}\right) + t_3\cdot\left(2-\sqrt{2}\right) + t_4\cdot\left(3-\sqrt{6}\right) + \delta k\Big)$$

We substitute $k = t_1 + t_2 + t_3 + t_4$, and obtain the following inequality:

$$\begin{aligned}
|VC(G)| &\leq (0.67)t_1 + 8\delta k + t_2 + (1.62)t_3 + (1.8)t_4 + \left(\sqrt{2}+1\right)\cdot\left(\frac{t_1}{2} + \frac{t_2}{10} + \frac{t_3}{10} + \frac{3t_4}{50} + \delta k\right)\\
&= (1.88)t_1 + (1.25)t_2 + (1.87)t_3 + (1.95)t_4 + \left(\sqrt{2}+9\right)\delta k\\
&< (1.95)k + \left(\sqrt{2}+9\right)\delta k \qquad \text{(using } t_3 + t_4 + t_1 + t_2 = k\text{)}\\
&\leq (2-\varepsilon)k, \quad \text{for appropriately small constants } \varepsilon, \delta > 0
\end{aligned}$$

This proves the soundness condition and it completes the proof of Theorem 14. Note that the result holds under the Unique Games Conjecture. To prove the result in a weaker assumtion of $\mathsf{P} \neq \mathsf{NP}$, it would require to show that $|VC(G)| < (1.36)k$ instead of $|VC(G)| < (1.95)k$. That would require tighter analysis of the cost of k-median instances than the one done in this work.

In the next section, we extend the above techniques to give the bi-criteria inapproximability results for the Euclidean k-median and k-means problems.

4 Bi-criteria Hardness of Approximation

In the previous section, we showed that the k-median problem cannot be approximated to any factor smaller than $(1+\varepsilon)$, where ε is some positive constant. The next step in the *beyond worst-case* discussion is to study the bi-criteria approximation algorithms. That is, we allow the algorithm to choose more than k centers and analyse whether it produces a solution that is close to the optimal solution with respect to k centers? Since the algorithm is allowed to output more than k centers we can hope to get a better approximate solution. An interesting question in this regard would be: *Does there exist a PTAS (polynomial time approximation scheme) for the k-median/k-means problem when the algorithm is allowed to choose βk centers for some constant $\beta > 1$?* In other words, is there an $(1+\varepsilon, \beta)$-approximation algorithm? Note that here we compare the cost of βk centers with the optimal cost with respect to k centers. See Definition 3 in Section 1 for formal definition of (α, β) bi-criteria approximation algorithms.

In this section, we show that even with βk centers, the k-means/k-median problems cannot be approximated within any factor smaller than $(1+\varepsilon')$, for some constant $\varepsilon' > 0$. The following theorem state this result formally.

▶ **Theorem 21** (k-median). *For any constant $1 < \beta < 1.015$, there exists a constant $\varepsilon > 0$ such that there is no $(1+\varepsilon, \beta)$-approximation algorithm for the Euclidean k-median problem assuming the Unique Games Conjecture.*

▶ **Theorem 22** (k-means). *For any constant $1 < \beta < 1.28$, there exists a constant $\varepsilon > 0$ such that there is no $(1+\varepsilon, \beta)$-approximation algorithm for the Euclidean k-means problem assuming the Unique Games Conjecture. Moreover, the same result holds for any $1 < \beta < 1.1$ under the assumption that* NP $\neq$ NP.

First, let us prove the bi-criteria inapproximability result for the k-median problem.

4.1 Bi-criteria Inapproximability: k-Median

In this subsection, we give a proof of Theorem 21. Let us define a few notations. Suppose $\mathcal{I} = (\mathcal{X}, k)$ be some k-median instance. Then, $OPT(\mathcal{X}, k)$ denote the optimal k-median cost of $\mathcal{X}$. Similarly, $OPT(\mathcal{X}, \beta k)$ denote the optimal βk-median cost of $\mathcal{X}$ (or the optimal cost of $\mathcal{X}$ with βk centers). We use the same reduction as we used in the previous section for showing the hardness of approximation of the k-median problem. Based on the reduction, we establish the following theorem.

▶ **Theorem 23.** *There is an efficient reduction from* Vertex Cover *on bounded degree triangle-free graphs G (with m edges) to Euclidean k-median instances $\mathcal{I} = (\mathcal{X}, k)$ that satisfies the following properties:*

1. *If G has a vertex cover of size k, then $OPT(\mathcal{X}, k) \leq m - k/2$*
2. *For any constant $1 < \beta < 1.015$, there exists constants $\varepsilon, \delta > 0$ such that if G has no vertex cover of size $\leq (2-\varepsilon) \cdot k$, then $OPT(\mathcal{X}, \beta k) \geq m - k/2 + \delta k$.*

Proof. Since the reduction is the same as we discussed in Section 1.2 and 3, we keep all notations the same as before. Also, note that Property 1 in this theorem is the same as Property 1 of Theorem 14. Therefore, the proof is also the same as we did in Section 3.1. Now, we directly move to the proof of Property 2.

The proof is almost the same as we gave in Section 3.2. However, it has some minor differences since we consider the optimal cost with respect to βk centers instead of k centers. Now, we prove the following contrapositive statement: "For any constants $1 < \beta < 1.015$ and $\varepsilon > 0$, there exists constants $\varepsilon, \delta > 0$ such that if $OPT(\mathcal{X}, \beta k) < (m - k/2 + \delta k)$ then G has a vertex cover of size at most $(2-\varepsilon)k$". Let $\mathcal{C}$ denote an optimal clustering of $\mathcal{X}$ with βk centers. We classify its optimal clusters into two categories: (1) *star* and (2) *non-star*. Further, we sub-classify the star clusters into the following two sub-categories:

(a) Clusters composed of exactly one edge. Let these clusters be: $P_1, P_2, \ldots, P_{t_1}$.
(b) Clusters composed of at least two edges. Let these clusters be: $S_1, S_2, \ldots, S_{t_2}$.

Similarly, we sub-classify the non-star clusters into the following two sub-categories:

(i) Clusters with a maximum matching of size two. Let these clusters be: $W_1, W_2, \ldots, W_{t_3}$
(ii) Clusters with a maximum matching of size at least three. Let these clusters be: $Y_1, Y_2, \ldots, Y_{t_4}$

Note that $t_1 + t_2 + t_3 + t_4$ equals βk. Suppose, we first compute a vertex cover of all the clusters except the single edge clusters: $P_1, \ldots, P_{t_1}$. Let that vertex cover be VC'. Now, some vertices in VC' might also cover the edges in $P_1, \ldots, P_{t_1}$. Suppose there are t_1' single edge clusters

that remain uncovered by VC'. Without loss of generality, we assume that these clusters are $P_1, \ldots, P_{t_1'}$. By Lemma 20, we can cover these clusters with $(\frac{2t_1'}{3} + 8\delta k) \leq (\frac{2t_1}{3} + 8\delta k)$ vertices; otherwise the graph would have a vertex cover of size at most $(2k - \delta k)$, and the proof of Property 2 would be complete. Now, we bound the vertex cover of the entire graph in the following manner.

$$\begin{aligned}
|VC(G)| &\leq \sum_{i=1}^{t_1} |VC(P_i)| + \sum_{i=1}^{t_2} |VC(S_i)| + \sum_{i=1}^{t_3} |VC(W_i)| + \sum_{i=1}^{t_4} |VC(Y_i)| \\
&\leq \left(\frac{2t_1}{3} + 8\delta k\right) + t_2 + \sum_{i=1}^{t_3} \left(\left(\sqrt{2}+1\right)\delta(W_i) + 1.62\right) + \sum_{i=1}^{t_4} \left(\left(\sqrt{2}+1\right)\delta(Y_i) + 1.8\right), \\
&\qquad\qquad \text{(using Lemmas 18, 19, and 20)} \\
&= (0.67)t_1 + 8\delta k + t_2 + (1.62)t_3 + (1.8)t_4 + \left(\sqrt{2}+1\right)\left(\sum_{i=1}^{t_3}\delta(W_i) + \sum_{i=1}^{t_4}\delta(Y_i)\right)
\end{aligned}$$

Since the optimal cost $OPT(\mathcal{X}, \beta k) = \sum_{j=1}^{\beta k}\sqrt{m_j(m_j-1)} + \sum_{i=1}^{t_3}\delta(W_i) + \sum_{i=1}^{t_4}\delta(Y_i) \leq m - k/2 + \delta k$, we get $\sum_{i=1}^{t_3}\delta(W_i) + \sum_{i=1}^{t_4}\delta(Y_i) \leq m - k/2 + \delta k - \sum_{j=1}^{\beta k}\sqrt{m_j(m_j-1)}$. We substitute this value in the previous equation, and get the following inequality:

$$|VC(G)| \leq (0.67)t_1 + 8\delta k + t_2 + (1.62)t_3 + (1.8)t_4 + \left(\sqrt{2}+1\right)\cdot\left(m - k/2 - \sum_{j=1}^{\beta k}\sqrt{m_j(m_j-1)} + \delta k\right)$$

Using Lemma 10, we obtain the following inequalities:

1. For P_j, $\sqrt{m(P_j)\,(m(P_j)-1)} \geq m(P_j) - 1$ since $m(P_j) = 1$
2. For S_j, $\sqrt{m(S_j)\,(m(S_j)-1)} \geq m(S_j) - (2-\sqrt{2})$ since $m(S_j) \geq 2$
3. For W_j, $\sqrt{m(W_j)\,(m(W_j)-1)} \geq m(W_j) - (2-\sqrt{2})$ since $m(W_j) \geq 2$
4. For Y_j, $\sqrt{m(Y_j)\,(m(Y_j)-1)} \geq m(Y_j) - (3-\sqrt{6})$ since $m(Y_j) \geq 3$

We substitute these values in the previous equation, and get the following inequality:

$$\begin{aligned}
|VC(G)| \leq (0.67)t_1 + 8\delta k + t_2 + (1.62)t_3 + (1.8)t_4 + \left(\sqrt{2}+1\right)\cdot\Bigg(m - k/2 - \sum_{j=1}^{t_1}(m(P_j)-1) + \\
- \sum_{j=1}^{t_2}\left(m(S_j) - (2-\sqrt{2})\right) - \sum_{j=1}^{t_3}\left(m(W_j) - (2-\sqrt{2})\right) - \sum_{j=1}^{t_4}\left(m(Y_j) - (3-\sqrt{6})\right) + \delta k\Bigg)
\end{aligned}$$

Since $m = \sum_{j=1}^{t_1} m(P_j) + \sum_{j=1}^{t_2} m(S_j) + \sum_{j=1}^{t_3} m(W_j) + \sum_{j=1}^{t_4} m(Y_j)$, we get the following inequality:

$$\begin{aligned}
|VC(G)| &\leq (0.67)t_1 + 8\delta k + t_2 + (1.62)t_3 + (1.8)t_4 + \left(\sqrt{2}+1\right)\cdot\Bigg(-k/2 + t_1 + t_2\cdot\left(2-\sqrt{2}\right) + \\
&\qquad + t_3\cdot\left(2-\sqrt{2}\right) + t_4\cdot\left(3-\sqrt{6}\right) + \delta k\Bigg) \\
&= (0.67)t_1 + 8\delta k + t_2 + (1.62)t_3 + (1.8)t_4 + \left(\sqrt{2}+1\right)\cdot\Bigg(\frac{(\beta-1)k}{2} - \frac{\beta k}{2} + t_1 + t_2\cdot\left(2-\sqrt{2}\right) + \\
&\qquad + t_3\cdot\left(2-\sqrt{2}\right) + t_4\cdot\left(3-\sqrt{6}\right) + \delta k\Bigg)
\end{aligned}$$

Now, we substitute $\beta k = t_1 + t_2 + t_3 + t_4$, and obtain the following inequality:

$$
\begin{aligned}
|VC(G)| &\leq (0.67)t_1 + 8\delta k + t_2 + (1.62)t_3 + (1.8)t_4 + \left(\sqrt{2}+1\right)\cdot\left(\frac{(\beta-1)k}{2} + \frac{t_1}{2} + \frac{t_2}{10} + \frac{t_3}{10} + \frac{3t_4}{50} + \delta k\right) \\
&= (1.88)t_1 + (1.25)t_2 + (1.87)t_3 + (1.95)t_4 + \left(\sqrt{2}+1\right)\cdot\frac{(\beta-1)k}{2} + \left(\sqrt{2}+9\right)\delta k \\
&< (1.95)\beta k + \left(\sqrt{2}+1\right)\cdot\frac{(\beta-1)k}{2} + \left(\sqrt{2}+9\right)\delta k \qquad \text{(using } t_3+t_4+t_1+t_2=\beta k\text{)} \\
&< (3.16)\beta k - (1.21)k + \left(\sqrt{2}+9\right)\delta k \\
&\leq (2-\varepsilon)k, \quad \text{for any } \beta < 1.015 \text{ and appropriately small constants } \varepsilon, \delta > 0
\end{aligned}
$$

This proves Property 2 and it completes the proof of Theorem 23. ◀

The following corollary states the main bi-criteria inapproximability result for the k-median problem.

▶ **Corollary 24.** *There exists a constant $\varepsilon' > 0$ such that for any constant $1 < \beta < 1.015$, there is no $(1+\varepsilon', \beta)$-approximation algorithm for the k-median problem assuming the Unique Games Conjecture.*

Proof. In the proof of Corollary 15, we showed that $k \geq \frac{m}{2\Delta}$ for all the hard Vertex Cover instances. Therefore, the second property of Theorem 23, implies that $OPT(\mathcal{X}, \beta k) \geq (m - \frac{k}{2}) + \delta k \geq (1 + \frac{\delta}{2\Delta}) \cdot (m - \frac{k}{2})$. Thus, the k-median problem can not be approximated within any factor smaller than $1 + \frac{\delta}{2\Delta} = 1 + \Omega(\varepsilon)$, with βk centers for any $\beta < 1.015$. ◀

The proof for the bi-criteria inapproximability of the k-means problem works in a similar manner. We deffer its proof to Appendix B.

References

1. Ankit Aggarwal, Amit Deshpande, and Ravi Kannan. Adaptive sampling for k-means clustering. In Irit Dinur, Klaus Jansen, Joseph Naor, and José D. P. Rolim, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, 12th International Workshop, APPROX 2009, and 13th International Workshop, RANDOM 2009, Berkeley, CA, USA, August 21-23, 2009. Proceedings*, volume 5687 of *Lecture Notes in Computer Science*, pages 15–28. Springer, 2009. doi:10.1007/978-3-642-03685-9_2.
2. S. Ahmadian, A. Norouzi-Fard, O. Svensson, and J. Ward. Better guarantees for k-means and euclidean k-median by primal-dual algorithms. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 61–72, October 2017. doi:10.1109/FOCS.2017.15.
3. Nir Ailon, Ragesh Jaiswal, and Claire Monteleoni. Streaming k-means approximation. In Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, and A. Culotta, editors, *Advances in Neural Information Processing Systems 22*, pages 10–18. Curran Associates, Inc., 2009. URL: http://papers.nips.cc/paper/3812-streaming-k-means-approximation.pdf.
4. Daniel Aloise, Amit Deshpande, Pierre Hansen, and Preyas Popat. Np-hardness of euclidean sum-of-squares clustering. *Mach. Learn.*, 75(2):245–248, May 2009. doi:10.1007/s10994-009-5103-0.
5. Vijay Arya, Naveen Garg, Rohit Khandekar, Adam Meyerson, Kamesh Munagala, and Vinayaka Pandit. Local search heuristics for k-median and facility location problems. *SIAM Journal on Computing*, 33(3):544–562, 2004. doi:10.1137/S0097539702416402.
6. P. Austrin, S. Khot, and M. Safra. Inapproximability of vertex cover and independent set in bounded degree graphs. In *2009 24th Annual IEEE Conference on Computational Complexity*, pages 74–80, 2009. doi:10.1109/CCC.2009.38.

7 Pranjal Awasthi, Moses Charikar, Ravishankar Krishnaswamy, and Ali Kemal Sinop. The Hardness of Approximation of Euclidean k-Means. In Lars Arge and János Pach, editors, *31st International Symposium on Computational Geometry (SoCG 2015)*, volume 34 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 754–767, Dagstuhl, Germany, 2015. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/LIPIcs.SOCG.2015.754.

8 Sayan Bandyapadhyay and Kasturi Varadarajan. On Variants of k-means Clustering. In Sándor Fekete and Anna Lubiw, editors, *32nd International Symposium on Computational Geometry (SoCG 2016)*, volume 51 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 14:1–14:15, Dagstuhl, Germany, 2016. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/LIPIcs.SoCG.2016.14.

9 Anup Bhattacharya, Dishant Goyal, and Ragesh Jaiswal. Hardness of approximation of euclidean k-median. *CoRR*, abs/2011.04221, 2020. arXiv:2011.04221.

10 Mihai Bundefineddoiu, Sariel Har-Peled, and Piotr Indyk. Approximate clustering via core-sets. In *Proceedings of the Thiry-Fourth Annual ACM Symposium on Theory of Computing*, STOC '02, page 250–257, New York, NY, USA, 2002. Association for Computing Machinery. doi:10.1145/509907.509947.

11 Jarosław Byrka, Thomas Pensyl, Bartosz Rybicki, Aravind Srinivasan, and Khoa Trinh. An improved approximation for k-median and positive correlation in budgeted optimization. *ACM Trans. Algorithms*, 13(2), 2017. doi:10.1145/2981561.

12 Ramaswamy Chandrasekaran and Arie Tamir. Open questions concerning weiszfeld's algorithm for the fermat-weber location problem. *Math. Program.*, 44(1-3):293–295, 1989. doi:10.1007/BF01587094.

13 Moses Charikar, Sudipto Guha, Éva Tardos, and David B. Shmoys. A constant-factor approximation algorithm for the k-median problem (extended abstract). In *Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing*, STOC '99, page 1–10, New York, NY, USA, 1999. Association for Computing Machinery. doi:10.1145/301250.301257.

14 Ke Chen. On coresets for k-median and k-means clustering in metric and euclidean spaces and their applications. *SIAM Journal on Computing*, 39(3):923–947, 2009. doi:10.1137/070699007.

15 Michael B. Cohen, Yin Tat Lee, Gary Miller, Jakub Pachocki, and Aaron Sidford. *Geometric Median in Nearly Linear Time*, page 9–21. Association for Computing Machinery, New York, NY, USA, 2016. doi:10.1145/2897518.2897647.

16 Vincent Cohen-Addad. A fast approximation scheme for low-dimensional k-means. In Artur Czumaj, editor, *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2018, New Orleans, LA, USA, January 7-10, 2018*, pages 430–440. SIAM, 2018. doi:10.1137/1.9781611975031.29.

17 Vincent Cohen-Addad and Karthik C. S. Inapproximability of clustering in lp metrics. In David Zuckerman, editor, *60th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2019, Baltimore, Maryland, USA, November 9-12, 2019*, pages 519–539. IEEE Computer Society, 2019. doi:10.1109/FOCS.2019.00040.

18 Vincent Cohen-Addad, Karthik C. S., and Euiwoong Lee. On approximability of clustering problems without candidate centers. In Dániel Marx, editor, *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms, SODA 2021, Virtual Conference, January 10 - 13, 2021*, pages 2635–2648. SIAM, 2021. doi:10.1137/1.9781611976465.156.

19 Vincent Cohen-Addad, Philip N. Klein, and Claire 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)*, 00:353–364, 2016. doi:doi.ieeecomputersociety.org/10.1109/FOCS.2016.46.

20 Sanjoy Dasgupta. The hardness of k-means clustering. Technical Report CS2008-0916, Department of Computer Science and Engineering, University of California San Diego, 2008.

21 Zvi Drezner and Horst W Hamacher. *Facility location: applications and theory*. Springer Science & Business Media, 2001.

22 Dan Feldman, Morteza Monemizadeh, and Christian Sohler. A PTAS for k-means clustering based on weak coresets. In *Proceedings of the twenty-third annual symposium on Computational geometry*, SCG '07, pages 11–18, New York, NY, USA, 2007. ACM. doi:10.1145/1247069.1247072.

23 Zachary Friggstad, Mohsen Rezapour, and Mohammad R. Salavatipour. Local search yields a PTAS for k-means in doubling metrics. *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, 00:365–374, 2016. doi:doi.ieeecomputersociety.org/10.1109/FOCS.2016.47.

24 Sudipto Guha and Samir Khuller. Greedy strikes back: Improved facility location algorithms. *J. Algorithms*, 31(1):228–248, 1999. doi:10.1006/jagm.1998.0993.

25 Ragesh Jaiswal, Amit Kumar, and Sandeep Sen. A simple D^2-sampling based PTAS for k-means and other clustering problems. *Algorithmica*, 70(1):22–46, 2014. doi:10.1007/s00453-013-9833-9.

26 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. In *Proceedings of the Eighteenth Annual Symposium on Computational Geometry*, SCG '02, page 10–18, New York, NY, USA, 2002. Association for Computing Machinery. doi:10.1145/513400.513402.

27 J. KRARUP and S. VAJDA. On torricelli's geometrical solution to a problem of fermat. *IMA Journal of Management Mathematics*, 8(3):215–224, 1997. doi:10.1093/imaman/8.3.215.

28 Amit Kumar, Yogish Sabharwal, and Sandeep Sen. Linear-time approximation schemes for clustering problems in any dimensions. *J. ACM*, 57(2):5:1–5:32, 2010. doi:10.1145/1667053.1667054.

29 Euiwoong Lee, Melanie Schmidt, and John Wright. Improved and simplified inapproximability for k-means. *Information Processing Letters*, 120:40–43, 2017. doi:10.1016/j.ipl.2016.11.009.

30 Shi Li. Approximating capacitated k-median with $(1 + \epsilon)k$ open facilities. In Robert Krauthgamer, editor, *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016*, pages 786–796. SIAM, 2016. doi:10.1137/1.9781611974331.ch56.

31 Shi Li. On uniform capacitated k-median beyond the natural lp relaxation. *ACM Trans. Algorithms*, 13(2), 2017. doi:10.1145/2983633.

32 Shi Li and Ola Svensson. Approximating k-median via pseudo-approximation. In *Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing*, STOC '13, page 901–910, New York, NY, USA, 2013. Association for Computing Machinery. doi:10.1145/2488608.2488723.

33 Meena Mahajan, Prajakta Nimbhorkar, and Kasturi Varadarajan. The planar k-means problem is np-hard. *Theoretical Computer Science*, 442:13–21, 2012. Special Issue on the Workshop on Algorithms and Computation (WALCOM 2009). doi:10.1016/j.tcs.2010.05.034.

34 Konstantin Makarychev, Yury Makarychev, and Ilya Razenshteyn. Performance of johnson-lindenstrauss transform for k-means and k-medians clustering. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2019, page 1027–1038, New York, NY, USA, 2019. Association for Computing Machinery. doi:10.1145/3313276.3316350.

35 Konstantin Makarychev, Yury Makarychev, Maxim Sviridenko, and Justin Ward. A Bi-Criteria Approximation Algorithm for k-Means. In Klaus Jansen, Claire Mathieu, José D. P. Rolim, and Chris Umans, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2016)*, volume 60 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 14:1–14:20, Dagstuhl, Germany, 2016. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/LIPIcs.APPROX-RANDOM.2016.14.

36 Jirí Matousek. On approximate geometric k-clustering. *Discret. Comput. Geom.*, 24(1):61–84, 2000. doi:10.1007/s004540010019.

37 Nimrod Megiddo and Kenneth J. Supowit. On the complexity of some common geometric location problems. *SIAM Journal on Computing*, 13(1):182–196, 1984. doi:10.1137/0213014.

38 P. Milasevic and G. R. Ducharme. Uniqueness of the spatial median. *Ann. Statist.*, 15(3):1332–1333, September 1987. doi:10.1214/aos/1176350511.

39 Stanislav Minsker. Geometric median and robust estimation in Banach spaces. *Bernoulli*, 21(4):2308–2335, 2015. doi:10.3150/14-BEJ645.

40 Andrea Vattani. The hardness of k-means clustering in the plane. Technical report, Department of Computer Science and Engineering, University of California San Diego, 2009.

41 Dennis Wei. A constant-factor bi-criteria approximation guarantee for k-means++. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems 29*, pages 604–612. Curran Associates, Inc., 2016. URL: http://papers.nips.cc/paper/6309-a-constant-factor-bi-criteria-approximation-guarantee-for-k-means.pdf.

42 E. WEISZFELD. Sur le point pour lequel la somme des distances de n points donnes est minimum. *Tohoku Mathematical Journal, First Series*, 43:355–386, 1937.

A Proof of Lemma 9

▶ **Lemma 25.** *Let $A = \{a_1, \ldots, a_n\}$ and $B = \{b_1, \ldots, b_n\}$ be any two sets of n points in $\mathbb{R}^d$. If the pairwise distances between points within A is the same as pairwise distance between points within B. That is, for all $i, j \in \{1, \ldots, n\}$, $\|a_i - a_j\| = \|b_i - b_j\|$. Then the optimal 1-median cost of A is the same as the optimal 1-median cost of B.*

Let $co(A)$ and $co(B)$ denote the convex hulls of A and B, respectively. We split the proof of Lemma 25 in two parts. In the first part (Lemma 26), we show that there exists a distance preserving transformation $\mathcal{R}$ from $co(A)$ to $co(B)$ such that $\mathcal{R}(a_i) = b_i$ for every $i \in \{1, \ldots, n\}$. By distance preserving transformation, we mean that for any two points $x, y \in co(A)$, the distance $\|x - y\|$ is preserved after applying the transformation $\mathcal{R}$, i.e., $\|x - y\| = \|\mathcal{R}(x) - \mathcal{R}(y)\|$. In the second part (Lemma 27), we show that applying the transformation $\mathcal{R}$ preserves the optimal 1-median cost of A.

▶ **Lemma 26.** *Given two sets of points $A = \{a_1, a_2, \ldots, a_n\}$ and $B = \{b_1, b_2, \ldots, b_n\}$ in $\mathbb{R}^d$ such that $\|a_i - a_j\| = \|b_i - b_j\|$ for all $i, j \in \{1, \ldots, n\}$. Then there exists a distance preserving transformation $\mathcal{R}\colon co(A) \to co(B)$ such that $\mathcal{R}(a_i) = b_i$ for every $i \in \{1, \ldots, n\}$.*

Proof. Let $\boldsymbol{X_i}$ be a vector[4] defined as $\boldsymbol{a_i} - \boldsymbol{a_1}$ for every $\boldsymbol{a_i} \in A$. Similarly, we define a vector $\boldsymbol{Y_i} := \boldsymbol{b_i} - \boldsymbol{b_1}$ for every $\boldsymbol{b_i} \in B$. We will use these vectors to define the transformation $\mathcal{R}$. For now, note the following property of inner product of $\boldsymbol{X_i}$ and $\boldsymbol{X_j}$.

$$\langle \boldsymbol{X_i}, \boldsymbol{X_j} \rangle = \langle \boldsymbol{Y_i}, \boldsymbol{Y_j} \rangle \quad \text{for every } i, j \in \{1, \ldots, n\} \tag{1}$$

The proof of the above property follows from the following sequence of inequalities:

$$\begin{aligned}
2 \cdot \langle \boldsymbol{X_i}, \boldsymbol{X_j} \rangle &= \|\boldsymbol{X_i}\|^2 + \|\boldsymbol{X_j}\|^2 - \|\boldsymbol{X_i} - \boldsymbol{X_j}\|^2 \\
&= \|\boldsymbol{Y_i}\|^2 + \|\boldsymbol{Y_j}\|^2 - \|\boldsymbol{X_i} - \boldsymbol{X_j}\|^2, && \because \|\boldsymbol{X_i}\| = \|\boldsymbol{a_i} - \boldsymbol{a_1}\| = \|\boldsymbol{b_i} - \boldsymbol{b_1}\| = \|\boldsymbol{Y_i}\| \\
& && \quad \text{for every } 1 \le i \le n \\
&= \|\boldsymbol{Y_i}\|^2 + \|\boldsymbol{Y_j}\|^2 - \|\boldsymbol{Y_i} - \boldsymbol{Y_j}\|^2, && \because \|\boldsymbol{X_i} - \boldsymbol{X_j}\| = \|\boldsymbol{a_i} - \boldsymbol{a_j}\| = \|\boldsymbol{b_i} - \boldsymbol{b_j}\| = \|\boldsymbol{Y_i} - \boldsymbol{Y_j}\| \\
&= 2 \cdot \langle \boldsymbol{Y_i}, \boldsymbol{Y_j} \rangle
\end{aligned}$$

In other words, the triangles (a_1, a_i, a_j) and (b_1, b_i, b_j) are congruent for all $i, j \in \{1, \ldots, n\}$. Therefore, the inner product $\langle \boldsymbol{X_i}, \boldsymbol{X_j} \rangle$ is the same as $\langle \boldsymbol{Y_i}, \boldsymbol{Y_j} \rangle$.

[4] For better readability, we boldfaced the vector symbols to distinguish them from any scalar quantity.

Now, we describe the transformation $\mathcal{R}$ from $co(A)$ to $co(B)$. By the definition of $co(A)$, any point $\boldsymbol{x} \in co(A)$ can be expressed in the form $\sum_{i=1}^{n} \lambda_i \cdot \boldsymbol{a_i}$ for some $0 \leq \lambda_i's \leq 1$ and $\sum_{i=1}^{n} \lambda_i = 1$. Equivalently, $\boldsymbol{x}$ can be expressed as $\boldsymbol{a_1} + \sum_{i=2}^{n} \lambda_i \cdot \boldsymbol{X_i}$. For $\boldsymbol{x} \in co(A)$, we define the transformation $\mathcal{R}$ as $\mathcal{R}(\boldsymbol{x}) \coloneqq \sum_{i=1}^{n} \lambda_i \cdot \boldsymbol{b_i}$. Again, $\mathcal{R}(\boldsymbol{x})$ can be equivalently expressed as $\boldsymbol{b_1} + \sum_{i=2}^{n} \lambda_i \cdot \boldsymbol{Y_i}$. It is easy to see that $\lambda_i \cdot \boldsymbol{b_i}$ indeed belongs to $co(B)$ since $0 \leq \lambda_i \leq 1$ and $\sum_{i=1}^{n} \lambda_i = 1$. Now, we show that $\mathcal{R}$ is a distance preserving transformation. Let $\boldsymbol{x} \coloneqq \boldsymbol{a_1} + \sum_{i=2}^{n} \lambda_i \cdot \boldsymbol{X_i}$ and $\boldsymbol{y} \coloneqq \boldsymbol{a_1} + \sum_{i=2}^{n} \gamma_i \cdot \boldsymbol{X_i}$ be any two points in $co(A)$. The following sequence of inequalities prove that $\|\boldsymbol{x} - \boldsymbol{y}\| = \|\mathcal{R}(\boldsymbol{x}) - \mathcal{R}(\boldsymbol{y})\|$.

$$
\begin{aligned}
\|\boldsymbol{x} - \boldsymbol{y}\|^2 &= (\boldsymbol{x} - \boldsymbol{y})^T (\boldsymbol{x} - \boldsymbol{y}) \\
&= \left(\sum_{i=2}^{n} (\lambda_i - \gamma_i) \cdot \boldsymbol{X_i} \right)^T \cdot \left(\sum_{i=2}^{n} (\lambda_i - \gamma_i) \cdot \boldsymbol{X_i} \right) \\
&= \sum_{i=2}^{n} \sum_{j=2}^{n} (\lambda_i - \gamma_i) \cdot (\lambda_j - \gamma_j) \cdot \langle \boldsymbol{X_i}, \boldsymbol{X_j} \rangle \\
&= \sum_{i=2}^{n} \sum_{j=2}^{n} (\lambda_i - \gamma_i) \cdot (\lambda_j - \gamma_j) \cdot \langle \boldsymbol{Y_i}, \boldsymbol{Y_j} \rangle, \qquad \text{(using Equation 1)} \\
&= \left(\sum_{i=2}^{n} (\lambda_i - \gamma_i) \cdot \boldsymbol{Y_i} \right)^T \cdot \left(\sum_{i=2}^{n} (\lambda_i - \gamma_i) \cdot \boldsymbol{Y_i} \right) \\
&= (\mathcal{R}(\boldsymbol{x}) - \mathcal{R}(\boldsymbol{y}))^T (\mathcal{R}(\boldsymbol{x}) - \mathcal{R}(\boldsymbol{y})) \\
&= \|\mathcal{R}(\boldsymbol{x}) - \mathcal{R}(\boldsymbol{y})\|^2
\end{aligned}
$$

This proves that $\mathcal{R}$ is a distance preserving transformation from $co(A)$ to $co(B)$. Moreover, note that $\mathcal{R}$ is a bijective function. It is possible that a vector $\boldsymbol{x} \in co(A)$ has multiple forms, say $\sum_{i=1}^{n} \lambda_i \cdot \boldsymbol{a_i}$ and $\sum_{i=1}^{n} \Lambda_i \cdot \boldsymbol{a_i}$. Therefore, it appears that $\boldsymbol{x}$ maps to different vectors in $co(B)$. However, it always maps to the same vector. For the sake of contradiction, assume that $\boldsymbol{x}$ maps to two different vectors $\boldsymbol{p} \coloneqq \sum_{i=1}^{n} \lambda_i \cdot \boldsymbol{b_i}$ and $\boldsymbol{q} \coloneqq \sum_{i=1}^{n} \Lambda_i \cdot \boldsymbol{b_i}$ in $co(B)$. Then $\|\boldsymbol{p} - \boldsymbol{q}\| \neq 0$. It contradicts the fact that $\mathcal{R}$ is a distance preserving transformation. Similarly, we can show that any two different vectors $\boldsymbol{x}, \boldsymbol{y} \in co(A)$ can not map to the same vector in $co(B)$. This proves that $\mathcal{R}$ is a bijective function.

Furthermore, note that $\mathcal{R}(\boldsymbol{a_i}) = \boldsymbol{b_i}$ for every $i \in \{1, \ldots, n\}$. To see this, consider $\lambda_i = 1$ and $\lambda_j = 0$ for all $j \in \{1, \ldots, n\} \setminus \{i\}$. Then $\boldsymbol{a_i} = \sum_{j=1}^{n} \lambda_j \cdot \boldsymbol{a}_j$ and therefore $\mathcal{R}(\boldsymbol{a_i}) = \sum_{j=1}^{n} \lambda_j \cdot \boldsymbol{b}_j = \boldsymbol{b_j}$. This completes the proof of the lemma. ◀

Similar to $\mathcal{R}$, we can also define a distance preserving transformation $\mathcal{R}^{-1}$ from $co(B)$ to $co(A)$. The transformation $\mathcal{R}^{-1}$ is defined such that for any $\boldsymbol{x} = \sum_{i=1}^{n} \lambda_i \cdot \boldsymbol{b_i} \in co(B)$,

$\mathcal{R}^{-1}(\boldsymbol{x}) = \sum_{i=1}^{n} \lambda_i \cdot \boldsymbol{a_i} \in co(A)$. Furthermore, as per this definition of $\mathcal{R}^{-1}$, $\mathcal{R}^{-1}(\boldsymbol{b_i}) = \boldsymbol{a_i}$ for every $i \in \{1, \ldots, n\}$. Now, we show that applying the transformation $\mathcal{R}$ on A preserves the optimal 1-median cost of A.

▶ **Lemma 27.** *If there exists distance preserving transformations $\mathcal{R}\colon co(A) \to co(B)$ and $\mathcal{R}^{-1}\colon co(B) \to co(A)$ such that $\mathcal{R}(\boldsymbol{a_i}) = \boldsymbol{b_i}$ and $\mathcal{R}^{-1}(\boldsymbol{b_i}) = \boldsymbol{a_i}$ for every $i \in \{1, \ldots, n\}$. Then the optimal 1-median cost of A is the same as the optimal 1-median cost of B.*

Proof. Recall that 1-median cost of an instance A with respect to a center $\boldsymbol{c} \in \mathbb{R}^d$ is denoted by $\Phi(\boldsymbol{c}, A) \equiv \sum_{\boldsymbol{a_i} \in A} \|\boldsymbol{a_i} - \boldsymbol{c}\|$. Let $\boldsymbol{c_1^*}$ be the optimal 1-median of A. Furthermore, we can assume that $\boldsymbol{c_1^*} \in co(A)$ since the optimal 1-median lies in the convex hull of A (see *e.g.* Remark 2.1 in [39]). Similarly, let $\boldsymbol{c_2^*} \in co(B)$ be the optimal 1-median of B. Now, we show that $\Phi(\boldsymbol{c_1^*}, A) \geq \Phi(\boldsymbol{c_2^*}, B)$ and $\Phi(\boldsymbol{c_1^*}, A) \leq \Phi(\boldsymbol{c_2^*}, B)$ using the following sequence of inequalities:

$$\begin{aligned}
\Phi(\boldsymbol{c_1^*}, A) &= \sum_{\boldsymbol{a_i} \in A} \|\boldsymbol{a_i} - \boldsymbol{c_1^*}\| \\
&= \sum_{\boldsymbol{a_i} \in A} \|\mathcal{R}(\boldsymbol{a_i}) - \mathcal{R}(\boldsymbol{c_1^*})\|, && \because \mathcal{R} \text{ preserves the pairwise distances} \\
&= \sum_{\boldsymbol{b_i} \in B} \|\boldsymbol{b_i} - \mathcal{R}(\boldsymbol{c_1^*})\|, && \because \mathcal{R}(a_i) = b_i \\
&\geq \sum_{\boldsymbol{b_i} \in B} \|\boldsymbol{b_i} - \boldsymbol{c_2^*}\|, && \because \boldsymbol{c_2^*} \text{ is the optimal 1-median of } B \\
&= \Phi(\boldsymbol{c_2^*}, B)
\end{aligned}$$

Similarly, we show that $\Phi(\boldsymbol{c_2^*}, B) \geq \Phi(\boldsymbol{c_1^*}, A)$ as follows:

$$\begin{aligned}
\Phi(\boldsymbol{c_2^*}, B) &= \sum_{\boldsymbol{b_i} \in B} \|\boldsymbol{b_i} - \boldsymbol{c_2^*}\| \\
&= \sum_{\boldsymbol{b_i} \in B} \|\mathcal{R}^{-1}(\boldsymbol{b_i}) - \mathcal{R}^{-1}(\boldsymbol{c_2^*})\|, && \because \mathcal{R}^{-1} \text{ preserves the pairwise distances} \\
&= \sum_{\boldsymbol{a_i} \in A} \|\boldsymbol{a_i} - \mathcal{R}^{-1}(\boldsymbol{c_2^*})\|, && \because \mathcal{R}^{-1}(b_i) = a_i \\
&\geq \sum_{\boldsymbol{a_i} \in A} \|\boldsymbol{a_i} - \boldsymbol{c_1^*}\|, && \because \boldsymbol{c_1^*} \text{ is the optimal 1-median of } A \\
&= \Phi(\boldsymbol{c_1^*}, A)
\end{aligned}$$

This proves that $\Phi(\boldsymbol{c_1^*}, A) = \Phi(\boldsymbol{c_2^*}, B)$. Hence it proves the lemma. ◀

Therefore, Lemma 26 and 27 together proves Lemma 25.

B Bi-criteria Inapproximability: k-means

Here, we again use the same reduction that we used earlier for the k-median problem in Sections 1.2, 3, and 4.1. Using this, we establish the following theorem.

▶ **Theorem 28.** *There is an efficient reduction from* Vertex Cover *on bounded degree triangle-free graphs G (with m edges) to Euclidean k-means instances $\mathcal{I} = (\mathcal{X}, k)$ that satisfies the following properties:*

1. *If G has a vertex cover of size k, then $OPT(\mathcal{X}, k) \leq m - k$*
2. *For any $1 < \lambda \leq 2$ and $\beta < \frac{2}{7} \cdot \left(\lambda + \frac{5}{2}\right)$, there exists constants $\varepsilon, \delta > 0$ such that if G has no vertex cover of size $\leq (\lambda - \varepsilon) \cdot k$, then $OPT(\mathcal{X}, \beta k) \geq m - k + \delta k$.*

This theorem is simply an extension of the result of Awasthi *et al.* [7] to the bi-criteria setting. Now, let us prove this theorem.

B.1 Completeness

Note that the proof of completeness is already given in [7]. Therefore, we just describe the main components of the proof for the sake of clarity. To understand the proof, let us define some notations used in [7]. Suppose F is a subgraph of G. For a vertex $v \in V(F)$, let $d_F(v)$ denote the number of edges in F that are incident on v. Note that, the optimal center for 1-means problem is simply the centroid of the point set. Therefore, we can compute the optimal 1-means cost of F. The following lemma states the optimal 1-means cost of F.

▶ **Lemma 29** (Claim 4.3 [7]). *Let F be a subgraph of G with r edges. Then, the optimal 1-means cost of F is $\sum_v d_F(v) \left(1 - \frac{d_F(v)}{r}\right)$*

The following corollary bounds the optimal 1-means cost of a star cluster. This corollary is implicitly stated in the proof of Claim 4.4 of [7].

▶ **Corollary 30.** *The optimal 1-means cost of a star cluster with r edges is $r - 1$.*

Using the above corollary, we give the proof of completeness. Let $V = \{v_1, \ldots, v_k\}$ be a vertex cover of G. Let S_i denote the set of edges covered by v_i. If an edge is covered by two vertices i and j, then we arbitrarily keep the edge either in S_i or S_j. Let m_i denote the number of edges in S_i. We define $\{\mathcal{X}(S_1), \ldots, \mathcal{X}(S_k)\}$ as a clustering of the point set $\mathcal{X}$. Now, we show that the cost of this clustering is at most $m - k$. Note that each S_i forms a star graph with its edges sharing the common vertex v_i. The following sequence of inequalities bound the optimal k-means cost of $\mathcal{X}$.

$$OPT(\mathcal{X}, k) \leq \sum_{i=1}^{k} \Phi^*(S_i) \overset{(Corollary\ 30)}{=} \sum_{i=1}^{k} (m(S_i) - 1) = m - k.$$

B.2 Soundness

For the proof of soundness, we prove the following contrapositive statement: "For any constant $1 < \lambda \leq 2$ and $\beta < \frac{2}{7} \cdot \left(\lambda + \frac{5}{2}\right)$, there exists constants $\varepsilon, \delta > 0$ such that if $OPT(\beta k) \leq (m - k + \delta k)$ then G has a vertex cover of size at most $(\lambda - \varepsilon)k$, for $\varepsilon = \Omega(\delta)$." Let $\mathcal{C}$ denote an optimal clustering of $\mathcal{X}$ with βk centers. We classify its optimal clusters into two categories: (1) *star* and (2) *non-star*. Suppose there are t_1 star clusters: $S_1, \ldots, S_{t_1}$, and t_2 non-star clusters: $F_1, F_2, \ldots, F_{t_2}$. Note that $t_1 + t_2$ equals βk. The following lemma bounds the optimal 1-means cost of a non-star cluster.

▶ **Lemma 31** (Lemma 4.8 [7]). *The optimal 1-means cost of any non-star cluster F with m edges is at least $m - 1 + \delta(F)$, where $\delta(F) \geq \frac{2}{3}$. Furthermore, there is an edge $(u, v) \in E(F)$ such that $d_F(u) + d_F(v) \geq m + 1 - \delta(F)$.*

In the original statement of the lemma in [7], the authors mentioned a weak bound of $\delta(F) > 1/2$. However, in the proof of their lemma they have shown $\delta(F) > 2/3 > 1/2$. This difference does not matter when we consider inapproximability of the k-means problem. However, this difference improves the β value in bi-criteria inapproximability of the k-means problem.

▶ **Corollary 32** ([7]). *Any non-star cluster F has a vertex cover of size at most $1 + \frac{5}{2} \cdot \delta(F)$.*

Proof. Suppose (u, v) be an edge in F that satisfies the property: $d_F(u) + d_F(v) \geq m + 1 - \delta(F)$, by Lemma 31. This means that u and v covers at least $m(F) - \delta(F)$ edges of F. We pick u and v in the vertex cover, and for the remaining $\delta(F)$ edges we pick one vertex per edge. Therefore, F has a vertex cover of size at most $2 + \delta(F)$. Since $\delta(F) \geq \frac{2}{3}$, by Lemma 31, we get $2 + \delta(F) \leq 1 + \frac{5}{2} \cdot \delta(F)$. Hence, F has a vertex cover of size at most $1 + \frac{5}{2} \cdot \delta(F)$. This proves the corollary. ◀

Now, the following sequence of inequalities bound the vertex cover size of the enire graph G.

$$\begin{aligned}
|VC(G)| &\leq \sum_{i=1}^{t_1} |VC(S_i)| + \sum_{i=1}^{t_2} |VC(F_i)| \\
&\leq t_1 + \sum_{i=1}^{t_2} \left(1 + \frac{5}{2} \cdot \delta(F_i)\right) \quad \text{(using Corollary 32)} \\
&= t_1 + t_2 + \frac{5}{2} \cdot \sum_{i=1}^{t_2} \delta(F_i)
\end{aligned}$$

Since the optimal k-means cost $OPT(\mathcal{X}, \beta k) = \sum_{i=1}^{t_1} (m(S_i) - 1) + \sum_{i=1}^{t_2} (m(F_i) - 1 + \delta(F_i)) \leq m - k + \delta k$, and $t_1 + t_2 = \beta k$. Therefore, $\sum_{i=1}^{t_2} \delta(F_i) \leq (\beta - 1)k + \delta k$. On substituting this value in the previous equation, we get the following inequality:

$$\begin{aligned}
|VC(G)| &\leq t_1 + t_2 + \frac{5}{2} \cdot (\beta - 1)k + \frac{5}{2} \cdot \delta k \\
&= \beta k + \frac{5}{2} \cdot (\beta - 1)k + \frac{5}{2} \cdot \delta k, \quad (\because t_1 + t_2 = \beta k) \\
&\leq (\lambda - \varepsilon)k, \quad \text{for any } \beta < \frac{2}{7} \cdot \left(\lambda + \frac{5}{2}\right) \text{ and appropriately small constants } \varepsilon, \delta > 0
\end{aligned}$$

This proves the soundness condition and thus completes the proof of Theorem 28.

Next, we state a corollary of Theorem 28 that gives the main bi-criteria inapproximability result for the k-means problem.

▶ **Corollary 33.** *For any constant $1 < \beta < 1.28$, there exists a constant $\varepsilon' > 0$ such that there is no $(1 + \varepsilon', \beta)$-approximation algorithm for the k-means problem assuming the Unique Games Conjecture. Moreover, the same result holds for any $1 < \beta < 1.1$ under the assumption that $\mathsf{P} \neq \mathsf{NP}$.*

Proof. Suppose Vertex Cover can not be approximated to any factor smaller than $\lambda - \varepsilon$, for some constants $\varepsilon, \lambda > 0$. In the proof of Corollary 15, we showed that $k \geq \frac{m}{2\Delta}$ for all the hard Vertex Cover instances. In that case, the second property of Theorem 28 implies that $OPT(\mathcal{X}, \beta k) \geq (m - k) + \delta k \geq (1 + \frac{\delta}{2\Delta}) \cdot (m - k)$. Thus, the k-means problem can not be approximated within any factor smaller than $1 + \frac{\delta}{2\Delta} = 1 + \Omega(\varepsilon)$, with βk centers. Now, let us compute the value of β using the value of λ. We know that $\beta < \frac{2}{7} \cdot \left(\lambda + \frac{5}{2}\right)$. Consider the following two cases:

- By Corollary 13, Vertex Cover is hard to approximate within any factor smaller than $2 - \varepsilon$ on bounded degree triangle-free graphs assuming UGC. Hence $\lambda = 2$ and thus $\beta < 1.28$ assuming UGC.

- By Theorem 7, Vertex Cover is hard to approximate within any factor smaller than 1.36 on bounded degree triangle-free graphs assuming $\mathsf{P} \neq \mathsf{NP}$. Hence $\lambda = 1.36$ and thus $\beta < 1.1$ assuming $\mathsf{P} \neq \mathsf{NP}$.

This completes the proof of the corollary. ◀