2025 IEEE 66th Annual Symposium on Foundations of Computer Science (FOCS) | 979-8-3315-7132-0/25/$31.00 ©2025 IEEE | DOI: 10.1109/FOCS63196.2025.00016

An Improved Greedy Approximation for (Metric) k-Means

Moses Charikar
Stanford University
moses@cs.stanford.edu

Vincent Cohen-Addad
Google Research
cohenaddad@google.com

Ruiquan Gao
Stanford University
ruiquan@cs.stanford.edu

Fabrizio Grandoni
IDSIA, USI-SUPSI
fabrizio.grandoni@gmail.com

Euiwoong Lee
University of Michigan
euiwoong@umich.edu

Ernest van Wijland
Université Paris-Cité
ernest.vanwijland@irif.fr

***Abstract*—Clustering is a basic task in data analysis and machine learning, and the optimization of clustering objectives are well-studied optimization problems; amongst these, the k-Means objective is arguably the most well known. Given a collection of points in a metric space, the goal is to partition them into k clusters, each with an associated center, so as to minimize the sum of squared distances of points to their cluster centers. In this paper, we present a polynomial-time $3 + 2\sqrt{2} + \varepsilon < 5.83$-approximation algorithm for k-Means in general metrics. This substantially improves on the current-best $(9 + \varepsilon)$-approximation in [Ahmadian, Norouzi-Fard, Svensson, Ward - FOCS'17, SICOMP'20], and even slightly improves on the 5.92-approximation in [Cohen-Addad, Esfandiari, Mirrokni, Narayanan - STOC'22] for the Euclidean special case.**

A natural approach for k-Means is to leverage Lagrangian Multiplier Preserving (LMP) approximations for the facility location problem. The previous best results for k-Means build upon an adaptation of an LMP 3-approximation for facility location with metric connection costs in [Jain, Vazirani - J.ACM'01] based on a primal-dual method, rather than on the improved LMP greedy 2-approximation for the same problem in [Jain, Mahdian, Markakis, Saberi, Vazirani - J.ACM'03]. The barrier to using the improved LMP algorithm was that no adaptation of this algorithm and its analysis to the case of squared metric connection costs was known (since squared distances violate triangle inequality). Our main contribution is overcoming this barrier by providing such an adaptation. This new LMP approximation algorithm is then combined with the framework recently introduced in [Cohen-Addad, Grandoni, Lee, Schwiegelshohn, Svensson - STOC'25] for the related (metric) k-Median problem.

***Index Terms*—Approximation Algorithm, Clustering, k-Means**

I. Introduction

In a generic clustering problem we are given a collection of points together with a dissimilarity measure (i.e. distance function) between pairs of points. The high-level goal is to partition the points into a "small" number of *clusters* so that similar points are clustered together while dissimilar ones are clustered separately. One of the most fundamental and best-studied clustering problems is *k-Means*. Here we are given a collection D of n points (or *clients*) and a collection F of *centers* (or *facilities*), as well as a integer $k > 0$. We are also given metric distances $d : (D \cup F) \times (D \cup F) \to \mathbb{R}_{\geq 0}$. Our goal is to select a set S of k centers (the *open* centers) so as to minimize the sum of the squared distances from each client to the closest open center[1], i.e.,

$$\sum_{j \in D} d^2(j, S) = \sum_{j \in D} \min_{i \in S} d^2(j, i).$$

Observe that each feasible solution S naturally induces k clusters where the cluster associated with $i \in S$ is given by the clients that are closer to i than to any other center in S (breaking ties arbitrarily). We next use OPT_k to denote a reference optimal solution and opt_k to denote its cost. When k is clear from the context, we simply use OPT and opt.

k-Means is NP-hard and well-studied in terms of approximation algorithms. It is impossible to approximate it (in polynomial time) better than a factor $1 + 8/e \approx 3.94$ [1]. The current-best (polynomial-time) approximation factor for this problem is $9+\varepsilon$ for any constant $\varepsilon > 0$ by Ahmadian, Norouzi-Fard, Svensson, and Ward [2]. Their algorithm is based on a primal-dual Lagrangian Multiplier Preserving (LMP) 9-approximation algorithm for facility location with squared metric connection costs (similar in spirit to a classical primal-dual 3-approximation algorithm by Jain and Vazirani [3] for metric connection costs), plus a careful way to combine so-called bi-point solutions that introduces a $1 + \varepsilon$ factor only in the approximation (rather than a larger constant factor as in prior work). This improved on an earlier 25-approximation by Gupta and Tangwongsan [4] based on local search: we will need this result later. Interestingly a $(1 + 8/e + \varepsilon)$-approximation can be obtained in FPT time [5].

Very often in practice one considers the Euclidean special case of k-Means (which in the literature is often just called k-Means), where the clients are n points in the (d-dimensional)

R.G. was supported by a Stanford Graduate Fellowship. F.G. was partially supported by the SNF Grants 200021-200731 and 200021-236706. E.L. was supported in part by NSF 2236669. E.v.W. was supported in part by the French PEPR integrated projects EPIQ (ANR-22-PETQ-0007).

[1]It can equivalently be phrased as the problem of partitioning the given points (D) into k disjoint clusters and finding the best center (amongst F) for each cluster, to minimize the k-Means objective.

979-8-3315-7132-0/25/$31.00 ©2025 IEEE
DOI 10.1109/FOCS63196.2025.00016

Euclidean space, with the respective distances. In this case one is allowed to select any point in the Euclidean space as a center. However known reductions [6]–[8] allow one to focus on a set of candidate centers F of size $n^{O_\varepsilon(1)}$ while introducing a factor $1+\varepsilon$ in the approximation for any constant $\varepsilon > 0$. Therefore one can (essentially) see the Euclidean case as a special case of the metric one. The best-known approximation factor for the Euclidean case is 5.92 by Cohen-Addad, Esfandiari, Mirrokni, Narayanan [9]. This improves on a sequence of increasingly better approximations for the problem: a $(9+\varepsilon)$-approximation by Kanungo, Mount, Netanyahu, Piatko, Silverman, and Wu [10], a 6.36-approximation by Ahmadian et al. [2], and a 6.13-approximation by Grandoni, Ostrovsky, Rabani, Schulman, and Venkat [11].

Our main result is a 5.83-approximation for (metric) k-Means. More precisely, we get the following:

Theorem 1. *For any constant $\varepsilon > 0$, there is a polynomial-time randomized $(3 + 2\sqrt{2} + \varepsilon)$-approximation for (metric) k-Means.*

We remark that the above result not only substantially improves the current best $(9+\varepsilon)$-approximation for the general metric case [2], but even slightly the current best 5.92-approximation for the Euclidean case [9]. We overview our approach in Section II.

A. Related Work

k-Means belongs to the family of k-clustering problems where the target number of clusters k is fixed. Other famous examples are k-Center and k-Median. In k-Center one wishes to select a set S of k centers so as to minimize the maximum distance from any client to S, i.e., the objective function to minimize is $\max_{j\in D} d(j, S)$. This problem admits a simple greedy 2-approximation which is best possible unless $P = NP$ [12], [13]. k-Median is defined like k-Means, except that here one wishes to minimize the sum of the distances rather than squared distances, i.e., the objective function is $\sum_{j\in D} d(j, S)$. k-Median is very close to k-Means in terms of results and techniques. k-Median is hard to approximate below a factor $1 + 2/e$ [1]. For general metric distances, the first constant approximation was achieved by Charikar, Guga, Tardos and Shmoys [14]. After a very long sequence of improvements [15]–[21], the current-best $(2 + \varepsilon)$-approximation for this problem was very recently achieved by Cohen-Addad, Grandoni, Lee, Schwiegelshohn, and Svensson [22]. This is also the best result for the Euclidean case, improving on an earlier 2.406-approximation [9].

II. Overview of our Approach

At a very high level, our approach looks similar in spirit to the one leading to a recent $(2 + \varepsilon)$-approximation for the related k-Median problem by Cohen-Addad, Grandoni, Lee, Schwiegelshohn, and Svensson [22]. In more detail, we exploit a combination of two different algorithms. The first one is a bicriteria approximation algorithm with the desired approximation factor, which opens $O(\log n/\varepsilon^2)$ more centers than the k allowed ones.

Theorem 2. *For any constant $\varepsilon \in (0, 1/6)$, there is a polynomial-time algorithm for k-Means that returns a solution containing at most $k+O(\log n/\varepsilon^2)$ centers and of cost at most $(3 + 2\sqrt{2} + \varepsilon)$opt.*

The second algorithm is a radically different $(5 + \varepsilon)$-approximation algorithm that only works for instances which are *stable* in the following sense. We say that an instance of k-Means is β-stable if $\text{opt}_{k-1} \geq (1 + \beta)\text{opt}_k$, where opt_{k-1} is the optimal cost for the same instance but with the target number of centers being $k - 1$.

Theorem 3. *For any constants $\varepsilon, \zeta > 0$, there is a polynomial-time randomized algorithm for k-Means that with high probability returns a solution of cost at most $(5 + \varepsilon)$opt assuming that the input instance is $(\zeta/\log n)$-stable.*

It is relatively easy to derive Theorem 1 from the above two theorems.

Proof of Theorem 1. We compute a set of feasible solutions, and return the cheapest one. Let Δ be the maximum extra number of centers computed by the algorithm from Theorem 2 (this number is independent from k). Let $k' = \max\{1, k-\Delta\}$. One solution is obtained by computing the optimum solution with one center if $k' = 1$, and otherwise by running the algorithm from Theorem 2 with target number of centers being k' (notice that it returns a solution with at most k centers, hence feasible). Furthermore, we run the algorithm from Theorem 3 for every integer $k'' \in (k', k]$ (thus obtaining solutions with $k'' \leq k$ centers, hence feasible).

If $\text{opt}_{k'}$ is not much larger than opt_k (more precisely $\text{opt}_{k'} \leq (1+\varepsilon)\text{opt}_k$), the first solution is $(1+\varepsilon)(3+2\sqrt{2}+\varepsilon)$ approximate. Otherwise, notice that there exists an integer $k'' \in (k', k]$ such that $\text{opt}_{k''} \leq (1 + \varepsilon)\text{opt}_k$ and $\text{opt}_{k''-1} \geq (1+\beta)\text{opt}_{k''}$ for $\beta \in \Omega(\varepsilon/\Delta) = \Omega(\varepsilon^3/\log n)$. For that value of k'' the corresponding instance is β-stable, hence the respective solution has cost at most $(5 + \varepsilon)\text{opt}_{k''} \leq (1 + \varepsilon)(5 + \varepsilon)\text{opt}_k$. The claim follows by rescaling ε by a constant factor. □

It remains to describe how the above two theorems are obtained. The algorithm from Theorem 3 is rather complex and its analysis is highly non-trivial (see Section 5 of the arXiv version). However, it is a relatively easy adaptation of the $(2 + \varepsilon)$-approximation algorithm in [22] for a similar notion of stable k-Median instances. The main difference is that we have to carefully use an approximate form of triangle inequality (see Lemmas 5 and 6). This also explains why we get a higher approximation factor $5 + \varepsilon$.

The main contribution of this paper is our proof of Theorem 2. In more detail, we consider the related facility location problem (with uniform opening costs). Recall that here, instead of a bound k on the number of centers, we are given a uniform facility cost f. We are now allowed to open an arbitrary number of centers/facilities S, and the objective function is

the total cost of the open facilities plus the squared distance from each client to the closest (open) facility in S, i.e.,

$$\text{cost}_{FL}(S) := f|S| + \sum_{j \in D} d^2(j, S)\,.$$

We say that an algorithm for the above problem is LMP Γ-approximate if it produces a feasible solution S such that

$$\Gamma \cdot f|S| + \sum_{j \in D} d^2(j, S) \leq \Gamma \cdot \text{opt}_{LP}(f). \qquad (1)$$

where $\text{opt}_{LP}(f) \leq \text{opt}$ is the optimal cost of a standard LP relaxation for the problem (see Section IV). In other words, the solution cost at most $\Gamma \cdot$opt even if we increase the facility cost of S by a factor Γ.

Recall that the current-best $(9+\varepsilon)$-approximation for (Metric) k-Means by Ahmadian et al. [2] builds upon a primal-dual LMP 9-approximation for facility location (with squared metric connection costs), which is inspired by a classical primal-dual 3-approximation by Jain and Vazirani [3] for the case of metric connection costs. Our main contribution is a *greedy* LMP $(3 + 2\sqrt{2})$-approximation for facility location with squared metric connection costs which is a non-trivial (but relatively simple) variant of the classical greedy LMP 2-approximation (JMMSV) for facility location with metric connection costs by Jain, Mahdian, Markakis, Saberi, and Vazirani [16]. It is instructive to recall how JMMSV works in order to see the differences.

a) The JMMSV Greedy Algorithm.: The algorithm has a variable α_j per client j (initialized to 0), a set S of open facilities (initialized to $\emptyset$), and a set A of *active* clients (initialized to D). At each point of time, the *bid* $bid(j,i)$ of client j towards facility i is $[\alpha_j - d(j,i)]^+$ if j is active and $[d(j,S) - d(j,i)]^+$ otherwise[2]. The variables α_j of active clients are increased uniformly until one of the following events happens:

(a) For some client j and $i \in S$, $\alpha_j \geq d(j,i)$. In that case j is removed from A and we say that j is connected to i;
(b) For some (not open) facility $i \notin S$, one has $\sum_{j \in D} bid(j,i) = f$.[3] In that case we open i, i.e., we add i to S. Furthermore, each $j \in A$ with $\alpha_j \geq d(j,i)$ is removed from A (and connected to j). Also, each inactive client $j \in D - A$ with $bid(j,i) > 0$ is reconnected to i.

b) Counterexample for the Naïve Extension.: The intuition behind the bids is as follows. On the one hand, the active clients j (which are not yet connected to an open facility) are willing to pay the difference $\alpha_j - d(j,i)$ (if positive) to open i, while reserving $d(j,i)$ for their own connection cost to i. On the other hand, the inactive clients $j \in D - A$ which are already connected to the closest facility in S, are willing to offer $d(j,S) - d(j,i)$ (if positive) towards the opening of facility i: if then i is actually opened, j reconnects to i which is closer (while altogether still spending α_j in total). This also motivates the term *greedy*. Notice that at any point of time active clients j satisfy $\alpha_j < d^2(j,S)$.

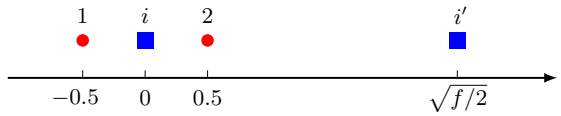

Fig. 1: Counterexample for the naïve variant of JMMSV algorithm. The example is on a line. Red circles represent the key clients, while blue squares represent the key facilities. The facility opening cost is f.

A naïve idea is to adapt the above algorithm to the case of squared distances by simply replacing $d(\cdot,\cdot)$ with $d^2(\cdot,\cdot)$. It is instructive to see why this attempt fails miserably (which also explains why prior work [2], [9] use variants of the Jain-Vazirani algorithm [3] instead, despite the worse approximation factor). Consider the example in Figure 1, where we have two facilities i and i', and two clients 1 and 2, all on a line, with distances specified in the figure. The facility opening cost is f. Consider the following (hypothetical but valid) execution of the algorithm.

- Time $f/2 + 0.25 - \varepsilon$: facility i' is open by client 2 and other clients not in the figure.
  - At that time, client 1 and 2's bid to i was $f/2 - \varepsilon$ each (so i was very close to be opened), but after the opening of i', client 2 lowers its bid to i to $(\sqrt{f/2} - 0.5)^2 - 0.5^2 = f/2 - \sqrt{f/2}$.
- Time $f/2 + \sqrt{f/2} + 0.25$: facility i is open as the bid from client 1 becomes $f/2 + \sqrt{f/2} + 0.25 - 0.25 = f/2 + \sqrt{f/2}$.
- To prove an LMP Γ-approximation, the dual constraint corresponding to a solution where 1 and 2 are connected to i (paying f for opening and $2 \cdot 0.5^2 = 0.5$ for connecting) requires that $\alpha_1 + \alpha_2 \leq f + 0.5\Gamma$. Our $\alpha_1 + \alpha_2$ is $f + \sqrt{f/2} + 0.5 - \varepsilon$, which is only an LMP $\Omega(\sqrt{f})$-approximation.

In this example, client 2 contributes to the opening of some center far away from it, with a contribution much less than its α value. Or equivalently, the connection cost $d^2(2,i')$ is strictly but only slightly smaller than α_2. This is the key (and essentially the only) reason why this naïve variant of JMMSV does not work; one can see that any other assumption between α_2 and $d^2(2,i')$ in the above example (realized by moving i' and clients not in the figure) would lead to a good LMP approximation as follows.

- If $d^2(2,i') = \alpha_2$, then 2's bid to i is $\alpha_2 - 0.25$ even after the opening of i', so the fact that i was open at time $\alpha_1 - \varepsilon$ (for infinitesimally small $\varepsilon > 0$) implies $\alpha_1 + \alpha_2 \leq f + 0.5$.
- If $d^2(2,i') \leq (1-\delta)\alpha_2$ for some constant $\delta > 0$, we could simply use an approximate version of the triangle

[2]We let $[a]^+ := \max\{0, a\}$.

[3]To be precise, this version of JMMSV establishes an equivalent notion of approximation, but not exactly the LMP 2-approximation defined in Equation (1); the latter is achieved when we run the algorithm with scaled opening cost $\hat{f} = 2f$, which our actual algorithms do with different scaling factors. In this overview, let us ignore this subtlety and conflate these two notions of approximations while just using f.

inequality to prove a good LMP approximation. For example, if $\delta = 1/2$, we would have $\alpha_1 \leq d^2(1,i') \leq 2d^2(2,i') + 4(d^2(1,i) + d^2(2,i)) \leq \alpha_2 + 2$. Since $\alpha_2 \leq f/2 + 0.25$ in this example, this would give us $\alpha_1 + \alpha_2 = f + O(1)$.

c) Our Algorithm.: This inspired us to construct a different variant of the JMMSV algorithm in the following spirit: if a client can contribute to the opening of a facility, its α value must be much higher than its connection cost to the facility (e.g., at least twice as high as it). In more detail, our LMP algorithm uses the following different bidding strategy. Let $\gamma > 1$ be a parameter to be fixed later. The bid of active clients is $[\alpha_j - \gamma d^2(j,i)]^+$. In particular notice that j cannot contribute to the opening of i if $\alpha_j < \gamma d^2(j,i)$ (while one might naturally expect it for $\alpha_j > d^2(j,i)$). However, whenever $\alpha_j \geq d^2(j,i)$ for some already open facility $i \in S$ and active clients j, we still connect j to i and make j inactive. In the above case, if $\gamma d^2(j,i) > \alpha_j$, we say that j is *indirectly connected* to i since j did not contribute to the opening of i with a positive bid. The inactive clients j that are connected to an open facility i whose opening they contributed to, are called *directly connected*. The bids for directly and indirectly connected clients in the later steps are different: they are $[\alpha_j - \gamma d^2(j,i)]^+$ for the indirectly connected ones and $[\gamma d^2(j,S) - \gamma d^2(j,i)]^+$ for the directly connected ones. When an indirectly connected client contributes to the opening of a facility, it becomes directly connected. For the above process, we are able to show that the final solution S satisfies:

$$\sum_{j \in D} \alpha_j \geq f|S| + \sum_{j \in D} d^2(j,S),$$

i.e., the cost of the solution is upper bounded by the sum of the variables α_j. Furthermore, we can establish the following inequality showing that α_j's, after scaling, form a dual-feasible solution.

$$\sum_{j \in D} \left[\alpha_j - \left(\gamma + 2 + \frac{2}{\gamma - 1} \right) d^2(j,i) \right]^+ \leq f, \quad \forall i \in F.$$

The above equations together imply that the overall algorithm is an LMP $(\gamma + 2 + \frac{2}{\gamma-1})$-approximation for the considered facility location problem. Fixing $\gamma = 1 + \sqrt{2}$ gives the claim.

d) Opening k Centers.: With the above LMP approximation for facility location at hand, we can now derive the result in Theorem 2 following an approach again close to [22]. A standard way to derive a ρ_{kMeans}-approximation algorithm for k-Means from an LMP ρ_{FL}-approximation for facility location with squared metric connection costs is to perform a binary search over the uniform facility cost f so as to obtain a solution with $k_1 < k$ open facilities for some facility cost f_1, and a solution with $k_2 > k$ open facilities for some facility cost $f_2 < f_1$, where the difference between f_1 and f_2 is very tiny. This is called a *bi-point* solution. Then the two solutions are combined together to obtain a solution opening k facilities. This combination is typically costly, hence leading to a factor ρ_{kMeans} substantially larger than ρ_{FL}[4]. We instead carefully modify the mentioned LMP $(3+2\sqrt{2})$-approximation for facility location (while increasing by ε the approximation factor) so that the variables α_j are increased in a logarithmic number of rounds only (with a small multiplicative increase at each round). Exploiting this fact and the walking-between-solutions framework in [22], we are able to construct a bi-point solution where $k_2 \leq k + O(\log n/\varepsilon^2)$. This leads to Theorem 2. Again, the details are rather technical but they are similar in spirit to [22].

III. Organisations

In this proceeding, we will only present our main contribution, the greedy algorithm. In the arXiv version, we present the full paper, including the bicriteria $(3+2\sqrt{2}+\varepsilon)$-approximation with $O(\log n/\varepsilon^2)$ extra centers (Section 5), which is built on the greedy algorithm, and the $(5+O(\sqrt{\varepsilon}))$-approximation for $(\frac{\zeta}{\log n})$-stable instances (Section 6).

IV. Preliminaries

a) LP Relaxations and Basic Assumptions: Given (D, F, d), the standard LP relaxations for k-Means and facility location with squared distances and uniform opening cost f are as follows.

$$\begin{aligned} \min \quad & \sum_{i \in F, j \in D} d(i,j)^2 x_{i,j} && (LP_{km}) \\ \text{s.t.} \quad & \sum_{i \in F} x_{i,j} \geq 1 && \forall j \in D \\ & y_i - x_{i,j} \geq 0 && \forall j \in D, i \in F \\ & \sum_{i \in F} y_i \leq k \\ & x, y \geq 0. \end{aligned}$$

$$\begin{aligned} \min \quad & \sum_{i \in F, j \in D} d(i,j)^2 x_{i,j} + f \cdot \sum_{i \in F} y_i && (LP_{FL}(f)) \\ \text{s.t.} \quad & \sum_{i \in F} x_{i,j} \geq 1 && \forall j \in D \\ & y_i - x_{i,j} \geq 0 && \forall j \in D, i \in F \\ & x, y \geq 0. \end{aligned}$$

Let $\text{opt}_{LP}(k)$ and $\text{opt}_{LP}(f)$ be the optimal values for LP_{km} and $LP_{FL}(f)$ respectively. Let $[a]^+ := \max(a, 0)$. The dual for $LP_{FL}(f)$ is:

$$\begin{aligned} \max \quad & \sum_{j \in D} \alpha_j && (DP_{FL}(f)) \\ \text{s.t.} \quad & \sum_{j \in D} [\alpha_j - d(i,j)^2]^+ \leq f && \forall i \in F \\ & \alpha \geq 0. \end{aligned}$$

[4]An exception is the approach in [2] which loses only a factor $(1+\varepsilon)$. However their approach is tailored to the specific LMP algorithms considered in that work and hard to adapt to others, as the authors openly admit.

Thanks to standard reductions, we can assume that distances are integers in $[1, n^3/\varepsilon]$ while loosing a factor $1+O(\varepsilon)$ in the approximation.

Lemma 4. *For any constants $\varepsilon > 0$ and $\alpha > 1$, given a polynomial-time α-approximation algorithm for k-Means on instances with distances in $\{1, \dots, n^3/\varepsilon\}$, there exists a polynomial-time $\alpha(1+O(\varepsilon))$-approximation algorithm for k-Means on general instances.*

Proof. Consider any input instance $(D \cup F, d)$ of k-Means. Assume that $n := |D|$ is large enough w.r.t. α, otherwise the problem can be solved by brute force in polynomial time. We guess $M := \max_{j \in D} d(j, \mathrm{OPT})$, where OPT is some optimal k-Means solution, by trying all the polynomially-many possibilities. If $M = 0$, the problem can be solved optimally in polynomial time, hence assume $M > 0$. Next, for each $(i,j) \in F \times D$, define $d'(j,i) = d'(i,j) = \max\{1, \lceil \frac{d(i,j)}{M} \frac{n}{\varepsilon} \rceil\}$ if $d(i,j) \le M$. Set all the remaining $d'(a,b)$, $a \neq b$, to $\frac{n^3}{\varepsilon}$. Finally, replace the $d'(i,j)$'s with the corresponding metric closure. We run the given α-approximation algorithm on the k-Means instance $(D \cup F, d')$, hence obtaining a solution S. It is easy to check that S is a good enough approximation for the input instance. □

b) Triangle Inequalities for Squared Distances.: Given three distances d, d_1, d_2 with $d \le d_1 + d_2$, the following lemma provides a useful upper bound on d^2 as a function of d_1 and d_2. The next lemma provides a similar upper bound in an analogous setting where $d \le d_1 + d_2 + d_3$.

Lemma 5. *Let $\gamma > 1$ be any constant. For any x, y we have $\gamma x^2 + \frac{\gamma}{\gamma-1} y^2 \ge (x+y)^2$.*

Proof. Let us show for which values of b the inequality $\gamma x^2 + by^2 \ge (x+y)^2$ holds. This is equivalent to requiring that the quadratic form $(\gamma-1)x^2 + (b-1)y^2 - 2xy$ is non negative. This happens iff $\gamma - 1 \ge 0$ (which is satisfied by assumption) and $(\gamma-1)(b-1) - 1 \ge 0$, which is satisfied for $b \ge \frac{\gamma}{\gamma-1}$. □

Lemma 6. *Let $\gamma > 1$ be any constant. For any x, y, z, we have $\gamma x^2 + \left(2 + \frac{2}{\gamma-1}\right)(y^2+z^2) \ge (x+y+z)^2$.*

Proof. The proof of this lemma consists of a rearrangement of the LHS and applying the inequality of arithmetic and geometric means:

$$\begin{aligned}
&\gamma \cdot x^2 + \left(2 + \frac{2}{\gamma-1}\right)(y^2+z^2) \\
&= x^2 + y^2 + z^2 + \left(\frac{\gamma-1}{2}x^2 + \frac{2}{\gamma-1}y^2\right) \\
&\qquad + \left(\frac{\gamma-1}{2}x^2 + \frac{2}{\gamma-1}z^2\right) + (y^2+z^2) \\
&\ge x^2 + y^2 + z^2 + 2xy + 2xz + 2yz \\
&\qquad\qquad \text{(AM-GM inequality)} \\
&= (x+y+z)^2. \qquad \square
\end{aligned}$$

V. Greedy Algorithm

We formally present the new greedy algorithm for facility location with squared distances outlined in Section II. Let $\gamma = 1 + \sqrt{2}$ and $\Gamma = \gamma + 2 + \frac{2}{\gamma-1} \approx 5.828$ be the desired approximation ratio. Given an instance (D, F, d, f) for facility location, let $\hat{f} = \Gamma f$. Algorithm 1 presents our new greedy algorithm achieving an LMP Γ-approximation.

We say that the clients in A are active, that the clients in IC are indirectly connected, and that the clients in DC are directly connected. The following standard lemma proves that these α values are enough to *pay* the connection costs and the (augmented) opening costs.

Lemma 7 (Approximation Guarantee)**.** *At the end of the execution of* Greedy Algorithm*, we have $\sum_{j \in D} \alpha_j \ge \sum_{j \in D} d^2(j,S) + |S|\hat{f}$.*

Proof. First, let us show that it is sufficient to prove that at the end of the execution, we have

$$\sum_{j \in DC} \alpha_j \ge \sum_{j \in DC} \gamma d^2(j,S) + \sum_{i \in S} \hat{f}\,. \tag{2}$$

Indeed, at the end of the execution, $A = \emptyset$, and for every $j \in IC$, $\alpha_j \ge d^2(j,S)$. Furthermore, $\gamma > 1$, therefore, (2) implies the lemma.

To this end, we prove that at any point in the execution of Greedy Algorithm, (2) holds.

The equality is initially true since $DC = \emptyset$ and $S = \emptyset$. Furthermore, since whenever the α-value of a client j reaches $d^2(j,S)$ it stops growing, no client is added to DC outside of when a facility is opened.

We now consider what happens when we open a facility i, i.e., add it to S. Let (α, S, A, IC, DC) be the state right before opening i, and set $DC' = \{j \in A \cup IC : \alpha_j \ge \gamma d^2(i,j)\}$, and $X = \{j \in DC : d^2(j,i) < d^2(j,S)\}$. The change of cost of the right-hand side of (2) is at most

$$\hat{f} + \sum_{j \in DC'} \gamma d^2(i,j) + \sum_{j \in X} \gamma (d^2(i,j) - d^2(j,S))\,.$$

Since the algorithm decided to open i, we also have

$$\hat{f} \le \sum_{j \in DC'} (\alpha_j - \gamma d^2(i,j)) + \sum_{j \in X} (\gamma d^2(j,S) - \gamma d^2(i,j))\,.$$

We thus get that the change of cost of the right-hand side is at most $\sum_{j \in DC'} \alpha_j$, which is the change of the left-hand side. □

Let $(\alpha_j)_{j \in D}$ be the final vector of α values. We prove the following lemma, which implies that α/Γ is a feasible solution for $DP_{FL}(f)$. Combined with Lemma 7, our solution S satisfies

$$\sum_{j \in D} d^2(j,S) + \Gamma |S| f \le \sum_{j \in D} \alpha_j \le \Gamma \cdot \mathrm{opt}_{LP}(f) \le \Gamma \cdot \mathrm{opt}_{FL}(f),$$

which finishes the proof of the LMP Γ-approximation.

Lemma 8 (Dual Feasibility)**.** *For every facility i, we have $\sum_{j \in D} [\alpha_j - \Gamma d^2(i,j)]^+ \le \hat{f}$.*

Algorithm 1 (GREEDY ALGORITHM)**.**

Initialization: Set $S \leftarrow \emptyset$ and $\alpha_j \leftarrow 0$ for every $j \in D$. Let $A \leftarrow D, IC \leftarrow \emptyset, DC \leftarrow \emptyset$

While $A \neq \emptyset$:

Increase the α-value of every $j \in A$ uniformly, until the following holds:

(1) There exists an unopened facility $i \notin S$ such that

$$\sum_{j \in A \cup IC} [\alpha_j - \gamma d^2(i,j)]^+ + \sum_{j \in DC} [\gamma d^2(j,S) - \gamma d^2(i,j)]^+ \geq \hat{f}.$$

If such event occurs, open i, i.e. add it to S.

(2) Some $j \in A$ has $\alpha_j \geq d^2(j,S)$.

Update A, IC, DC so that

- $A = \{j \in D : \alpha_j < d^2(j,S)\}$.
- $IC = \{j \in D : d^2(j,S) \leq \alpha_j < \gamma d^2(j,S)\}$.
- $DC = \{j \in D : \gamma d^2(j,S) \leq \alpha_j\}$.

Proof. Consider a facility i. Let $D^* = \{j \in D : \alpha_j > \Gamma d^2(i,j)\}$. Since $\alpha_j \leq \Gamma d^2(i,j)$ for any $j \notin D^*$, it remains to show that

$$\sum_{j \in D^*} \alpha_j \leq \hat{f} + \Gamma \cdot \sum_{j \in D^*} d^2(i,j) \ . \tag{3}$$

Let $s = |D^*|$ be the size of D^*. Without loss of generality, we assume that $D^* = [s]$ and it is sorted in non-decreasing order of α_j values, namely, $\alpha_1 \leq \ldots \leq \alpha_s$.

For each $j \in D^*$, let DC^j (resp., IC^j) be the set of clients j' in $[j-1]$ such that point $j' \in DC$ (resp., $j' \in IC \cup A$) at time $\alpha_j - \varepsilon$, where the value of $\varepsilon > 0$ can be an arbitrary positive number such that $\alpha_j - \varepsilon > \Gamma \cdot d^2(j,i)$ and such that no client becomes connected or no facility becomes open between $\alpha_j - \varepsilon$ (inclusive) and α_j (exclusive). Similarly, let S^j be the set S at time $\alpha_j - \varepsilon$. Then, for any $j \in D^*$ and $j' \in DC^j$, the squared distance between j and the facility i' that j' connects to at time $\alpha_j - \varepsilon$ should be strictly larger than $\alpha_j - \varepsilon$. This is because otherwise we will stop growing α_j at time $\alpha_j - \varepsilon$. Since ε here can be arbitrarily small, we have $d^2(j,i') \geq \alpha_j$. Further, applying Lemma 6, we have

$$\begin{aligned}
&\forall j \in [s], j' \in DC^j, \\
&\alpha_j \leq d^2(j,i') \leq (d(j,i) + d(j',i) + d(j',i'))^2 \\
&\quad \leq \gamma \cdot d^2(j',i') + \left(2 + \frac{2}{\gamma - 1}\right) \cdot (d^2(j,i) + d^2(j',i)) \\
&\quad \leq \gamma \cdot d^2(j',S^j) + \left(2 + \frac{2}{\gamma - 1}\right) \cdot (d^2(j,i) + d^2(j',i))
\end{aligned} \tag{4}$$

Note that $\alpha_j - \varepsilon > \Gamma \cdot d^2(j,i) > \gamma \cdot d^2(j,i)$ for any $j \in D^*$. At time $\alpha_j - \varepsilon$, facility i is unopened, because otherwise client j should be directly connected to i. Therefore, we have

$$\sum_{j' \in DC^j} \left[\gamma d^2(j',S^i) - \gamma d^2(j',i)\right]^+ + \sum_{j' \in IC^j} \left[\alpha_{j'} - \gamma d^2(j',i)\right]^+ + \sum_{j \leq j' \leq s} \left[\alpha_j - \varepsilon - \gamma d^2(j',i)\right]^+ < \hat{f}.$$

Again, since this ε can be arbitrarily small and $[v]^+ \geq v$ for any value of v, for any $j \in D^*$, we have the following inequality:

$$\sum_{j' \in DC^j} \left(\gamma d^2(j',S^i) - \gamma d^2(j',i)\right) + \sum_{j' \in IC^j} \left(\alpha_{j'} - \gamma d^2(j',i)\right) + \sum_{j \leq j' \leq s} \left(\alpha_j - \gamma d^2(j',i)\right) \leq \hat{f} \ .$$

By applying Eq. (4), we get

$$\sum_{j' \in DC^j} \alpha_j - \left(2 + \frac{2}{\gamma - 1}\right) \cdot \left(d^2(j,i) + d^2(j',i)\right) - \gamma d^2(j',i) + \sum_{j' \in IC^j} (\alpha_{j'} - \gamma d^2(j',i)) + \sum_{j \leq j' \leq s} (\alpha_j - \gamma d^2(j',i)) \leq \hat{f} \ ,$$

and we can simplify it as (using the definition of $\Gamma := \gamma + 2 + \frac{2}{\gamma - 1}$)

$$\begin{aligned}
&(s - j + 1 + |DC^j|) \cdot \alpha_j + \sum_{j' \in IC^j} \alpha_{j'} \\
&\leq \hat{f} + \gamma \sum_{j' \in D^*} d^2(j',i) + (\Gamma - \gamma) \sum_{j' \in DC^j} (d^2(j,i) + d^2(j',i)).
\end{aligned} \tag{β_j}$$

Next, we show that all Eq. (β_j) (for $j \in D^*$) together imply our objective Eq. (3). Let $A_{jj'}$ define the coefficient of $\alpha_{j'}$ in Eq. (β_j). That is, we have

$$A_{jj'} = \begin{cases} \mathbb{1}(j' \in IC^j) & \text{if } j' < j; \\ s - j + 1 + |DC^j| & \text{if } j' = j; \\ 0 & \text{if } j' > j. \end{cases}$$

And, we can rewrite Eq. (β_j) as

$$\sum_{j'\in[s]} A_{jj'}\alpha_{j'} \le \hat{f} + \gamma\cdot\sum_{j'\in[s]} d^2(j',i) + (\Gamma-\gamma)\cdot\sum_{j'<j}(1-A_{jj'})(d^2(j,i)+d^2(j',i)).$$

Let $\beta_1,\ldots,\beta_s$ be some coefficients such that $\sum_{j\in[s]}\beta_j\cdot\sum_{j'\in[s]}A_{jj'}\alpha_{j'} = \sum_{j\in[s]}\alpha_j$, where we view each α_j as a variable. As we have $A_{jj'}=0$ for any $j'>j$, we can easily compute each β_j one by one, using the values of $\beta_{j+1},\ldots,\beta_s$. According to the definition of $A_{j'j}$, we have

$$\beta_j = \frac{1}{A_{jj}}\cdot\left(1-\sum_{j'>j}\beta_{j'}A_{j'j}\right) . \tag{5}$$

Since $A_{j'j}\in\{0,1\}$ for any $j'>j$, to show $\beta_j\ge 0$, it suffices to show that $\sum_{j'>j}\beta_{j'}\le 1$. In fact, we will show the following stronger fact.

Fact 9. *Fix any $j\in\{0\}\cup[s]$. We have $\sum_{j'>j}\beta_{j'}\le\frac{s-j}{s-j+|DC^j|}$, where we define $DC^0:=\emptyset$.*

Proof. Observe that $A_{j'\ell}=0$ for any $j'<\ell\in[s]$. Since our definition of $\beta_{j'}$ ensures that the coefficient of each α_ℓ ($\ell\in[s]$) in $\sum_{j'\in[s]}\beta_{j'}\sum_{\ell\in[s]}A_{j'\ell}\alpha_\ell$ equals 1, i.e., $\sum_{j'\in[s]}\beta_{j'}A_{j'\ell}=1$ for any $\ell\in[s]$, this observation implies that the coefficient of each α_ℓ ($\ell>j$) is exactly 1 in $\sum_{j'>j}\beta_{j'}\sum_{\ell\in[s]}A_{j'\ell}\alpha_\ell$. That is,

$$\forall \ell>j,\qquad \sum_{j'>j}\beta_{j'}A_{j\ell}=1 .$$

Summing over all such $\ell>j$, we get

$$\sum_{j'>j}\left(\sum_{\ell>j}A_{j'\ell}\right)\beta_{j'} = s-j . \tag{6}$$

Recall the definition of $A_{j'\ell}$, where we have for any $j'>j$,

$$\begin{aligned} A_{j'j'} &= s-j'+1+|DC^{j'}| \\ &= s-j'+1+\sum_{j<\ell<j'}\mathbb{1}(\ell\in DC^{j'})+\sum_{\ell\le j}\mathbb{1}(\ell\in DC^{j'}), \\ A_{j'\ell} &= \mathbb{1}(\ell\in IC^{j'}) \qquad\qquad \forall j<\ell<j' . \end{aligned}$$

For any $j<\ell<j'$, we have either $\ell\in DC^{j'}$ or $\ell\in IC^{j'}$ (note that $IC^{j'}$ includes unconnected clients with index less than j' at time $\alpha_{j'}-\varepsilon$). For each $j'>j$, each $j<\ell<j'$ contributes exactly 1 to $\sum_{\ell>j}A_{j'\ell}$. Hence, for any $j'>j$, we have

$$\begin{aligned}\sum_{\ell>j}A_{j'\ell} &= s-j'+1+(j'-1)-j+\sum_{\ell\le j}\mathbb{1}(\ell\in DC^{j'}) \\ &\ge s-j+|DC^j| ,\end{aligned}$$

where the inequality results from the fact that $DC^{j'}\cap[j-1]\supseteq DC^j$ for $j'>j$. Applying this lower bound to Eq. (6), we get $s-j\ge(s-j+|DC^j|)\sum_{j'>j}\beta_{j'}$, which is equivalent to $\sum_{j'>j}\beta_{j'}\le\frac{s-j}{s-j+|DC^j|}$. □

Hence, we have $\beta_j\ge 0$ for any $j\in[s]$ according to our earlier discussions. Further, we have

$$\begin{aligned}\sum_{j\in[s]}\alpha_j &= \sum_{j\in[s]}\beta_j\cdot\left(\sum_{j'\in[s]}A_{jj'}\alpha_{j'}\right) \\ &\le \left(\hat{f}+\gamma\cdot\sum_{j\in[s]}d^2(j,i)\right)\cdot\sum_{j\in[s]}\beta_j \\ &\quad+(\Gamma-\gamma)\sum_{j'<j\in[s]}\beta_j(1-A_{jj'})(d^2(j,i)+d^2(j',i)) \\ &\le \hat{f}+\gamma\cdot\sum_{j\in[s]}d^2(j,i) \\ &\quad+(\Gamma-\gamma)\sum_{j'<j\in[s]}\beta_j(1-A_{jj'})(d^2(j,i)+d^2(j',i)).\end{aligned} \tag{7}$$

Finally, to finish the proof, we show that

$$\sum_{j\in[s]}\beta_j\cdot\sum_{j'<j}(1-A_{jj'})(d^2(j,i)+d^2(j',i)) \le \sum_{j\in[s]}d^2(j,i). \tag{8}$$

Rearranging the LHS of the inequality, we show that the LHS of Eq. (8) equals

$$\begin{aligned}&\sum_{j\in[s]}\sum_{j'<j}\beta_j\cdot(1-A_{jj'})\cdot d^2(j,i) \\ &\qquad+\sum_{j\in[s]}\sum_{j'<j}\beta_j\cdot(1-A_{jj'})\cdot d^2(j',i) \\ &=\sum_{j\in[s]}\left(\beta_j\sum_{j'<j}(1-A_{jj'})+\sum_{j'>j}\beta_{j'}(1-A_{j'j})\right)\cdot d^2(j,i) \\ &\le\sum_{j\in[s]}\left(\beta_j\cdot\sum_{j'<j}(1-A_{jj'})+\sum_{j'>j}\beta_{j'}\right)\cdot d^2(j,i) \\ &\qquad\qquad(\forall j'>j,\ A_{j'j}\in\{0,1\}) \\ &\le\sum_{j\in[s]}\left(\beta_j\cdot\sum_{j'<j}(1-A_{jj'})+\frac{s-j}{s-j+|DC^j|}\right)\cdot d^2(j,i) \\ &\qquad\qquad\text{(Fact 9)}\end{aligned}$$

According to the recurrence of β_j (Eq. (5)), we have $\beta_j\le\frac{1}{A_{jj}}=\frac{1}{s-j+1+|DC^j|}$. Note that, for any $j'<j$, $A_{jj'}=0$ if and only if $j'\notin IC^j$, i.e., $j'\in DC^j$. Therefore, $\sum_{j'<j}(1-A_{jj'})=|DC^j|$. We can further upper bound the LHS of Eq. (8) by

$$\begin{aligned}&\sum_{j\in[s]}\left(\frac{|DC^j|}{s-j+1+|DC^j|}+\frac{s-j}{s-j+|DC^j|}\right)\cdot d^2(j,i) \\ &\le\sum_{j\in[s]}\left(\frac{|DC^j|}{s-j+|DC^j|}+\frac{s-j}{s-j+|DC^j|}\right)\cdot d^2(j,i) \\ &=\sum_{j\in[s]}d^2(j,i)\end{aligned}$$

Putting Eqs. (7) and (8) together, we get

$$\sum_{j\in[s]} \alpha_j \le \hat{f} + \gamma \cdot \sum_{j\in[s]} d^2(j,i) + (\Gamma - \gamma) \cdot \sum_{j\in[s]} d^2(j,i)$$
$$= \hat{f} + \Gamma \cdot \sum_{i\in[s]} d^2(j,i) \ . \qquad \square$$

REFERENCES

[1] K. Jain, M. Mahdian, and A. Saberi, "A new greedy approach for facility location problems," in *Proceedings on 34th Annual ACM Symposium on Theory of Computing, May 19-21, 2002, Montréal, Québec, Canada*, 2002, pp. 731–740. [Online]. Available: http://doi.acm.org/10.1145/509907.510012

[2] S. Ahmadian, A. Norouzi-Fard, O. Svensson, and J. Ward, "Better guarantees for k-means and Euclidean k-median by primal-dual algorithms," *SIAM J. Comput.*, vol. 49, no. 4, 2020. [Online]. Available: https://doi.org/10.1137/18M1171321

[3] K. Jain and V. V. Vazirani, "Approximation algorithms for metric facility location and *k*-median problems using the primal-dual schema and Lagrangian relaxation," *J. ACM*, vol. 48, no. 2, pp. 274–296, 2001. [Online]. Available: http://doi.acm.org/10.1145/375827.375845

[4] A. Gupta and K. Tangwongsan, "Simpler analyses of local search algorithms for facility location," *CoRR*, vol. abs/0809.2554, 2008. [Online]. Available: http://arxiv.org/abs/0809.2554

[5] V. Cohen-Addad, A. Gupta, A. Kumar, E. Lee, and J. Li, "Tight FPT approximations for k-median and k-means," in *46th International Colloquium on Automata, Languages, and Programming, ICALP 2019, July 9-12, 2019, Patras, Greece*, ser. LIPIcs, C. Baier, I. Chatzigiannakis, P. Flocchini, and S. Leonardi, Eds., vol. 132. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019, pp. 42:1–42:14. [Online]. Available: https://doi.org/10.4230/LIPIcs.ICALP.2019.42

[6] J. Matoušek, "On approximate geometric k-clustering," *Discrete & Computational Geometry*, vol. 24, no. 1, pp. 61–84, 2000.

[7] W. F. De La Vega, M. Karpinski, C. Kenyon, and Y. Rabani, "Approximation schemes for clustering problems," in *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, 2003, pp. 50–58.

[8] D. Feldman, M. Monemizadeh, and C. Sohler, "A ptas for k-means clustering based on weak coresets," in *Proceedings of the twenty-third annual symposium on Computational geometry*, 2007, pp. 11–18.

[9] V. Cohen-Addad, H. Esfandiari, V. S. Mirrokni, and S. Narayanan, "Improved approximations for Euclidean *k*-means and *k*-median, via nested quasi-independent sets," in *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, S. Leonardi and A. Gupta, Eds. ACM, 2022, pp. 1621–1628. [Online]. Available: https://doi.org/10.1145/3519935.3520011

[10] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "A local search approximation algorithm for k-means clustering," *Comput. Geom.*, vol. 28, no. 2-3, pp. 89–112, 2004. [Online]. Available: https://doi.org/10.1016/j.comgeo.2004.03.003

[11] F. Grandoni, R. Ostrovsky, Y. Rabani, L. J. Schulman, and R. Venkat, "A refined approximation for euclidean k-means," *Inf. Process. Lett.*, vol. 176, p. 106251, 2022. [Online]. Available: https://doi.org/10.1016/j.ipl.2022.106251

[12] T. F. Gonzalez, "Clustering to minimize the maximum intercluster distance," *Theor. Comput. Sci.*, vol. 38, pp. 293–306, 1985. [Online]. Available: http://dx.doi.org/10.1016/0304-3975(85)90224-5

[13] D. S. Hochbaum and D. B. Shmoys, "A unified approach to approximation algorithms for bottleneck problems," *J. ACM*, vol. 33, no. 3, pp. 533–550, 1986. [Online]. Available: http://doi.acm.org/10.1145/5925.5933

[14] M. Charikar, S. Guha, É. Tardos, and D. 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, May 1-4, 1999, Atlanta, Georgia, USA*, J. S. Vitter, L. L. Larmore, and F. T. Leighton, Eds. ACM, 1999, pp. 1–10. [Online]. Available: https://doi.org/10.1145/301250.301257

[15] K. Jain and V. Vazirani, "Approximation algorithms for metric facility location and *k*-median problems using the primal-dual schema and Lagrangian relaxation," *J. ACM*, vol. 48, no. 2, pp. 274–296, 2001. [Online]. Available: http://doi.acm.org/10.1145/375827.375845

[16] K. Jain, M. Mahdian, E. Markakis, A. Saberi, and V. V. Vazirani, "Greedy facility location algorithms analyzed using dual fitting with factor-revealing LP," *J. ACM*, vol. 50, no. 6, pp. 795–824, 2003. [Online]. Available: https://doi.org/10.1145/950620.950621

[17] V. Arya, N. Garg, R. Khandekar, A. Meyerson, K. Munagala, and V. Pandit, "Local search heuristics for k-median and facility location problems," *SIAM J. Comput.*, vol. 33, no. 3, pp. 544–562, 2004. [Online]. Available: https://doi.org/10.1137/S0097539702416402

[18] S. Li and O. Svensson, "Approximating k-median via pseudo-approximation," *SIAM J. Comput.*, vol. 45, no. 2, pp. 530–547, 2016. [Online]. Available: https://doi.org/10.1137/130938645

[19] J. Byrka, T. W. Pensyl, B. Rybicki, A. Srinivasan, and K. Trinh, "An improved approximation for *k*-median and positive correlation in budgeted optimization," *ACM Trans. Algorithms*, vol. 13, no. 2, pp. 23:1–23:31, 2017. [Online]. Available: https://doi.org/10.1145/2981561

[20] V. Cohen-Addad, F. Grandoni, E. Lee, and C. Schwiegelshohn, "Breaching the 2 LMP approximation barrier for facility location with applications to *k*-median," in *Proceedings of the 2023 ACM-SIAM Symposium on Discrete Algorithms, SODA 2023, Florence, Italy, January 22-25, 2023*, N. Bansal and V. Nagarajan, Eds. SIAM, 2023, pp. 940–986. [Online]. Available: https://doi.org/10.1137/1.9781611977554.ch37

[21] K. N. Gowda, T. W. Pensyl, A. Srinivasan, and K. Trinh, "Improved bi-point rounding algorithms and a golden barrier for *k*-median," in *Proceedings of the 2023 ACM-SIAM Symposium on Discrete Algorithms, SODA 2023, Florence, Italy, January 22-25, 2023*, N. Bansal and V. Nagarajan, Eds. SIAM, 2023, pp. 987–1011. [Online]. Available: https://doi.org/10.1137/1.9781611977554.ch38

[22] V. Cohen-Addad, F. Grandoni, E. Lee, C. Schwiegelshohn, and O. Svensson, "A (2+ε)-approximation algorithm for metric k-median," in *Proceedings of the 57th ACM Symposium on Theory of Computing, STOC 2025*, 2025.