



Approximation Algorithms for the Lower-Bounded k -Median and Its Generalizations

Lu Han¹, Chunlin Hao², Chenchen Wu^{3(✉)}, and Zhenning Zhang²

¹ Academy of Mathematics and Systems Science, Chinese Academy of Sciences,
Beijing 100190, People's Republic of China

hanlu@amss.ac.cn

² Department of Operations Research and Information Engineering,
Beijing University of Technology, Beijing 100124, People's Republic of China

{haochl, zhangzhenning}@bjut.edu.cn

³ College of Science, Tianjin University of Technology,
Tianjin 300384, People's Republic of China

wu_chenchen_tjut@163.com

Abstract. In this paper, we consider the lower-bounded k -median problem (LB k -median) that extends the classical k -median problem. In the LB k -median, a set of facilities, a set of clients and an integer k are given. Every facility has its own lower bound on the minimum number of clients that must be connected to the facility if it is opened. Every facility-client pair has its connection cost. We want to open at most k facilities and connect every client to some opened facility, such that the total connection cost is minimized.

As our main contribution, we study the LB k -median and present our main bi-criteria approximation algorithm, which, for any given constant $\alpha \in [0, 1)$, outputs a solution that satisfies the lower bound constraints by a factor of α and has an approximation ratio of $\frac{1+\alpha}{1-\alpha}\rho$, where ρ is the state-of-art approximation ratio for the k -facility location problem (k -FL). Then, by extending the main algorithm to several general versions of the LB k -median, we show the versatility of our algorithm for the LB k -median. Last, through providing relationships between the constant α and the approximation ratios, we demonstrate the performances of all the algorithms for the LB k -median and its generalizations.

Keywords: k -median · Lower bounds · Approximation algorithm · Bi-criteria

1 Introduction

The uncapacitated facility location problem (UFL) has numerous applications in operations management and computer science. In this problem, we are given a set of facilities and a set of clients. Every facility has an associated opening cost, and every facility-client pair has an associated connection cost which is

proportional to the distance between the facility and client. The aim is to open some facilities and connect every client to an opened facility so as to minimize the total opening and connection cost. Since the UFL is a well-known NP-hard problem, researchers pay attention on designing approximation algorithms for it and its generalizations [3, 8, 10, 12, 15, 16, 18, 19]. For a minimization problem, a λ -approximation algorithm is a polynomial time algorithm which can output a solution for any instance of the problem, such that the cost of the solution is within a factor of λ of the cost of an optimal solution. For the UFL, under the assumption that the connection costs are metric (i.e., the connection costs are non-negative, symmetric and satisfy the triangle inequality), Li [12] presents the current best 1.488-approximation algorithm and assume that $P \neq NP$ Sviridenko [16] gives the 1.463-hardness of approximation.

However, in many real-life situations, the facility expects to be connected by a minimum number of clients for the profitable sake. In fact, the lower-bounded facility location problem (LBFL) characterizes these scenarios. Besides, the motivation of the lower bound constraints also comes from a data privacy perspective [1]. Compared with the UFL, in the LBFL, every facility is given an additional lower bound on the minimum number of clients that must be connected to the facility if it is opened. The goal is to find some facilities to open and connect every client to an opened facility without violating any lower bound constraints, such that the total opening as well as connection cost is minimized. The LBFL is introduced by Guha et al. [7] and Karger and Minkoff [11] simultaneously. Both give an $O(1)$ -bi-criteria approximation algorithm that approximately subjects to the lower bound constraints. For the special case of the LBFL where the lower bound of every facility is the same, by reducing the LBFL to the capacitated facility location problem (CFL), Svitkina [17] proposes the first true 448-approximation algorithm. Later, Ahmadian and Swamy [2] improve the approximation ratio to 82.6. For the general case of the LBFL where every facility has its own lower bound, Li [13] also reduces the LBFL to the CFL and gives the breakthrough true approximation algorithm which has a ratio of 4000.

When every facility in the UFL does not have an opening cost and the aim becomes to find at most k facilities to open and connect every client to some opened facility so as to minimize the sum of connection costs, we get the classical k -median problem. The k -median is another well-studied NP-hard problem and has various generalizations. [3–6, 9, 10, 14, 18, 19]. For the k -median, under the assumption that the connection costs are metric, Byrka et al. [5] present the state-of-art $(2.675 + \epsilon)$ -approximation algorithm and assume that $NP \not\subseteq DTIME(n^{O(\log \log n)})$. Jain et al. [9] offer the 1.736-hardness of approximation. Despite the fact that many meaningful and interesting general versions of the k -median have been considered in the literatures, to the best of our knowledge, very little work concentrates on studying the k -median with lower bounds. This situation stimulates us to pay attention on the lower-bounded k -median problem (LB k -median). Compared with the k -median, the LB k -median has extra lower bound constraints which need to be respected.

In this paper, we study the LB k -median and its generalizations. First, inspired by the previous works of Guha et al. [7] and Karger and Minkoff [11] on the LBFL, we propose our main $O(1)$ -bi-criteria approximation algorithm for the LB k -median, which satisfies the lower bound constraints by a factor of some given constant $\alpha \in [0, 1)$ and has an approximation ratio of $\frac{1+\alpha}{1-\alpha}\rho$, where ρ is the current best approximation ratio for the k -facility location problem (k -FL). The key idea behind the main algorithm relies on an observation that constructing and solving a new instance of the k -FL instead of the original instance of the LB k -median can easily obtain a solution respects the cardinality constraint, and then trying to guarantee every facility is connected by a certain amount of clients can give us a bi-criteria solution. Second, we extend the main algorithm to several generalizations of the LB k -median, including the lower-bounded k -facility location problem (LB k -FL), the lower-bounded knapsack median problem (LB knapsack median) and the prize-collecting lower-bounded k -median problem (PLB k -median). The algorithms for these generalizations involve constructing and solving new instances of the k -FL, the knapsack facility location problem (knapsack FL) and the prize-collecting k -facility location problem (P k -FL), respectively. Last but not least, we give the relationships between the given constant α and the approximation ratios of all the algorithms to demonstrate their performances. Particularly, we show that our algorithm for the LB k -median can give a nice approximation ratio while violating the lower bound constraints within an acceptable range.

The remainder of our paper is structured as follows. Section 2 presents our main $O(1)$ -bi-criteria approximation algorithm for the LB k -median. Section 3 extends the main algorithm to several general versions of the LB k -median. Section 4 demonstrates the performances of our algorithms. Due to space constraint, all proofs are removed but will further appear in a full version of this paper.

2 The Lower-Bounded k -median Problem

In this section, we present an $O(1)$ -bi-criteria approximation algorithm for the LB k -median. Subsection 2.1 describes the LB k -median and the relevant k -FL along with their integer programs. Subsection 2.2 presents our main algorithm for the LB k -median and its analysis.

2.1 Preliminaries for the LB k -median

In the LB k -median, we are given a set of facilities \mathcal{F} , a set of clients \mathcal{D} and an integer k . Every facility $i \in \mathcal{F}$ has an associated lower bound L_i on the minimum number of clients in \mathcal{D} that must be connected to the facility if it is opened. Every facility-client pair (i, j) , where $i \in \mathcal{F}$ and $j \in \mathcal{D}$, has an associated connection cost c_{ij} which is proportional to the distance between facility i and client j . Under the assumption that the connection costs are metric, the goal

is to find at most k facilities to open and connect every client to some opened facility, such that the total connection cost is minimized.

The LB k -median can be formulated as the following integer program:

$$\min \sum_{i \in \mathcal{F}} \sum_{j \in \mathcal{D}} c_{ij} x_{ij} \tag{1}$$

$$\text{s. t. } \sum_{i \in \mathcal{F}} x_{ij} \geq 1, \quad \forall j \in \mathcal{D}, \tag{2}$$

$$x_{ij} \leq y_i, \quad \forall i \in \mathcal{F}, j \in \mathcal{D}, \tag{3}$$

$$\sum_{j \in \mathcal{D}} x_{ij} \geq L_i y_i, \quad \forall i \in \mathcal{F}, \tag{4}$$

$$\sum_{i \in \mathcal{F}} y_i \leq k, \tag{5}$$

$$x_{ij} \in \{0, 1\}, \quad \forall i \in \mathcal{F}, j \in \mathcal{D}, \tag{6}$$

$$y_i \in \{0, 1\}, \quad \forall i \in \mathcal{F}. \tag{7}$$

In program (1–7), there are two types of variables ($\{x_{ij}\}_{i \in \mathcal{F}, j \in \mathcal{D}}, \{y_i\}_{i \in \mathcal{F}}$). The variable x_{ij} indicates whether client j is connected to facility i for any facility-client pair (i, j) where $i \in \mathcal{F}$ and $j \in \mathcal{D}$. The variable y_i indicates whether facility i is opened for any facility $i \in \mathcal{F}$. The objective function describes the total connection cost. The constraints (2) say that every client $j \in \mathcal{D}$ must be connected to some facility. The constraints (3) state that if a client j is connected to some facility $i \in \mathcal{F}$, then the facility must be opened. The constraints (4) guarantee that the lower bound of any opened facility cannot be violated. The constraint (5) shows that the number of opened facilities can not exceed k .

When every facility $i \in \mathcal{F}$ in the LB k -median has an associated opening cost f_i instead of the lower bound L_i and the aim becomes to find at most k facilities to open and connect every client to some opened facility so as to minimize the sum of opening costs as well as connection costs, we get the k -FL. By introducing the same variables ($\{x_{ij}\}_{i \in \mathcal{F}, j \in \mathcal{D}}, \{y_i\}_{i \in \mathcal{F}}$), as in the integer program (1–7), the k -FL can be formulated as the following integer program:

$$\min \sum_{i \in \mathcal{F}} f_i y_i + \sum_{i \in \mathcal{F}} \sum_{j \in \mathcal{D}} c_{ij} x_{ij} \tag{8}$$

$$\text{s. t. } \sum_{i \in \mathcal{F}} x_{ij} \geq 1, \quad \forall j \in \mathcal{D}, \tag{9}$$

$$x_{ij} \leq y_i, \quad \forall i \in \mathcal{F}, j \in \mathcal{D}, \tag{10}$$

$$\sum_{i \in \mathcal{F}} y_i \leq k, \tag{11}$$

$$x_{ij} \in \{0, 1\}, \quad \forall i \in \mathcal{F}, j \in \mathcal{D}, \tag{12}$$

$$y_i \in \{0, 1\}, \quad \forall i \in \mathcal{F}. \tag{13}$$

In program (8-13), the objective function consists of opening costs and connection costs.

Note that we have the following observation.

Lemma 1. *From program (1-7) and program (8-13), it is clear that with the same inputs of \mathcal{F} , \mathcal{D} and k , any feasible solution for the LB k -median is also a feasible solution for the k -FL.*

2.2 Algorithm for the LB k -median

For the LB k -median, we propose a bi-criteria approximation algorithm, that, for any given constant $\alpha \in [0, 1)$, outputs a solution in which every opened facility i is connected by at least αL_i clients and has a constant approximation ratio of $\frac{1+\alpha}{1-\alpha}\rho$, where ρ is the current best approximation ratio for the k -FL. Our main algorithm for the LB k -median consists of three steps. First of all, from the instance \mathcal{IN} of the LB k -median, we construct a new instance \mathcal{IN}' of the k -FL. Secondly, we apply existing approximation algorithm for the k -FL to solve the instance \mathcal{IN}' and obtain a solution (S', σ') where S' is the set of opened facilities and $\sigma' : \mathcal{D} \rightarrow S'$ is the corresponding connections of clients in \mathcal{D} to facilities in S' . Finally, we continually close some facility in S' and reconnect its clients to obtain a new solution (S, σ) which connects at least αL_i clients to every opened facility $i \in S$.

For any facility $i \in \mathcal{F}$, denote \mathcal{D}_i as the set of closest L_i clients to it in \mathcal{D} . Now we are ready to present our main algorithm.

Algorithm 1

Step 1 Construct a new instance of the k -FL.

For the instance $\mathcal{IN} = (\mathcal{F}, \mathcal{D}, k, \{L_i\}_{i \in \mathcal{F}}, \{c_{ij}\}_{i \in \mathcal{F}, j \in \mathcal{D}})$ of the LB k -median, pick a constant $\alpha \in [0, 1)$, get rid of the lower bounds $\{L_i\}_{i \in \mathcal{F}}$ from \mathcal{IN} and

$$\text{set } f_i := \frac{2\alpha}{1-\alpha} \sum_{j \in \mathcal{D}_i} c_{ij} \text{ for every } i \in \mathcal{F},$$

in order to obtain a new instance $\mathcal{IN}' = (\mathcal{F}, \mathcal{D}, k, \{f_i\}_{i \in \mathcal{F}}, \{c_{ij}\}_{i \in \mathcal{F}, j \in \mathcal{D}})$ of the k -FL.

Step 2 Solve the instance of the k -FL.

Solve new instance \mathcal{IN}' with the current best ρ -approximation algorithm for the k -FL (see [19]), where $\rho = 2 + \sqrt{3} + \epsilon$, and obtain a feasible solution (S', σ') , where S' is the set of opened facilities and $\sigma' : \mathcal{D} \rightarrow S'$ is a function that maps every client $j \in \mathcal{D}$ to the closest facility in S' . For any client $j \in \mathcal{D}$, let $\sigma'(j)$ denote its closest facility in S' .

Step 3 Construct a solution for the LB k -median.

Step 3.1 Initialization.

At the very beginning, set $S := S'$ and $\sigma(j) := \sigma'(j)$ for any $j \in \mathcal{D}$, define $l_i := |\{j \in \mathcal{D} : \sigma(j) = i\}|$ for any $i \in \mathcal{F}$ and $S_d := \{i \in S : l_i < \alpha L_i\}$.

Step 3.2 Close facilities and reconnect clients.

While $S_d \neq \emptyset$ **do**

Arbitrarily choose some facility $i \in S_d$ and close it. For every client j with $\sigma(j) = i$, reconnect it to its closest facility i' in $S \setminus \{i\}$ and update $\sigma(j) := i'$. Update $S := S \setminus \{i\}$. Update l_i for any facility $i \in \mathcal{F}$ and S_d .

Output solution (S, σ) .

Algorithm 1 provides a solution (S, σ) , where S is the set of opened facilities and $\sigma : \mathcal{D} \rightarrow S$ denotes the corresponding connections between clients in \mathcal{D} and facilities in S , for the LB k -median. For any client $j \in \mathcal{D}$, let $\sigma(j)$ denote the facility which is connected by j in solution (S, σ) .

The following theorem presents our main result for the LB k -median.

Theorem 1. *Algorithm 1 is a bi-criteria approximation algorithm for the LB k -median that produces a solution (S, σ) , which connects at least αL_i clients to every opened facility $i \in S$, and has an approximation ratio of $\frac{1+\alpha}{1-\alpha} \rho$ where α is a given constant in interval $[0, 1)$ and ρ is the current best approximation ratio of $2 + \sqrt{3} + \epsilon$ for the k -FL.*

Because of Step 3.2 in Algorithm 1, it is not hard to see that

$$|j \in \mathcal{D} : \sigma(j) = i| = l_i \geq \alpha L_i \text{ for any } i \in S,$$

which means the solution (S, σ) connects at least αL_i clients to every opened facility $i \in S$. The remainder of this section will put focus on analyzing the approximation ratio of our algorithm for the LB k -median.

Suppose that (S^*, σ^*) is the optimal solution for the instance \mathcal{IN} of the LB k -median, where S^* is the optimal set of opened facilities and $\sigma^* : \mathcal{D} \rightarrow S^*$ denotes the optimal corresponding connections. Let OPT_{lk} be the total cost of the solution (S^*, σ^*) for \mathcal{IN} , i.e., $OPT_{lk} = \sum_{j \in \mathcal{D}} c_{\sigma^*(j)j}$. For every client $j \in \mathcal{D}$, denote $\sigma^*(j)$ as the facility which is connected by j in solution (S^*, σ^*) . In order to provide the approximation ratio of Algorithm 1, the following lemmas are essential.

Lemma 2. *The total cost of the solution (S', σ') for the instance \mathcal{IN}' of the k -FL is within a factor of $\frac{1+\alpha}{1-\alpha} \rho$ of the total cost of the optimal solution (S^*, σ^*) for the instance \mathcal{IN} of the LB k -median, i.e.,*

$$\sum_{i \in S'} f_i + \sum_{j \in \mathcal{D}} c_{\sigma'(j)j} \leq \frac{1+\alpha}{1-\alpha} \rho \cdot OPT_{lk},$$

where $\alpha \in [0, 1)$ and $\rho = 2 + \sqrt{3} + \epsilon$.

Lemma 3. *The total cost of the solution (S, σ) for the instance \mathcal{IN} of the LB k -median is no more than the total cost of the solution (S', σ') for the instance \mathcal{IN}' of the k -FL, i.e.,*

$$\sum_{j \in \mathcal{D}} c_{\sigma(j)j} \leq \sum_{i \in S'} f_i + \sum_{j \in \mathcal{D}} c_{\sigma'(j)j}.$$

Integrating Lemma 2 with Lemma 3 implies the approximation ratio of Algorithm 1.

3 Generalizations of the Lower-Bounded k -median Problem

In this section, by extending our main algorithm to several more general versions of the LB k -median, we show the versatility of Algorithm 1. Subsection 3.1, 3.2 and 3.3 present algorithms for the LB k -FL, LB knapsack median and PLB k -median through altering only the first step, the first two steps and all the steps in Algorithm 1, respectively.

3.1 The Lower-Bounded k -facility Location Problem

Compared with the LB k -median, in the LB k -FL, every facility $i \in \mathcal{F}$ is given an additional opening cost f_i . The aim is to open at most k facilities and connect every client to some opened facility, such that the total opening and connection cost is minimized.

The algorithm for the LB k -FL is obtained by only modifying the first step in Algorithm 1 slightly.

Algorithm 2

Step 1 Construct a new instance of the k -FL.

For the instance $\mathcal{IN} = (\mathcal{F}, \mathcal{D}, k, \{L_i\}_{i \in \mathcal{F}}, \{f_i\}_{i \in \mathcal{F}}, \{c_{ij}\}_{i \in \mathcal{F}, j \in \mathcal{D}})$ of the LB k -FL, pick a constant $\alpha \in [0, 1)$, get rid of the lower bounds $\{L_i\}_{i \in \mathcal{F}}$ from \mathcal{IN} and

$$\text{set } f'_i := f_i + \frac{2\alpha}{1-\alpha} \sum_{j \in \mathcal{D}_i} c_{ij} \text{ for every } i \in \mathcal{F},$$

in order to obtain a new instance $\mathcal{IN}' = (\mathcal{F}, \mathcal{D}, k, \{f'_i\}_{i \in \mathcal{F}}, \{c_{ij}\}_{i \in \mathcal{F}, j \in \mathcal{D}})$ of the k -FL.

Step 2 Solve the instance of the k -FL.

Same as Step 2 in Algorithm 1. Solve new instance \mathcal{IN}' with the current best ρ -approximation algorithm for the k -FL (see [19]), where $\rho = 2 + \sqrt{3} + \epsilon$, and obtain a feasible solution (S', σ') .

Step 3 Construct a solution for the LB k -FL.

Same as Step 3 in Algorithm 1. At the end of this step, output solution (S, σ) .

The following theorem offers the result for the LB k -median.

Theorem 2. *Algorithm 2 is a bi-criteria approximation algorithm for the LB k -FL that produces a solution (S, σ) , which connects at least αL_i clients to every opened facility $i \in S$, and has an approximation ratio of $\frac{1+\alpha}{1-\alpha} \rho$ where $\alpha \in [0, 1)$ and $\rho = 2 + \sqrt{3} + \epsilon$.*

We skip the proof of this theorem since it is similar to the one for the LB k -median.

3.2 The Lower-Bounded Knapsack Median Problem

Compared with the LB k -median, in the LB knapsack median, the cardinality constraint is replaced with a knapsack constraint. More specifically, we are given some budget B instead of the integer k . Every facility $i \in \mathcal{F}$ has a weight w_i . We want to open a subset $S \subseteq \mathcal{F}$ of facilities which subjects to $\sum_{i \in S} w_i \leq B$, and connect every client to some opened facility, so as to minimize the total connection cost.

The algorithm for the LB knapsack median is offered by changing the first two steps in Algorithm 1.

Algorithm 3

Step 1 Construct a new instance of the knapsack FL.

For the instance $\mathcal{IN} = (\mathcal{F}, \mathcal{D}, B, \{L_i\}_{i \in \mathcal{F}}, \{w_i\}_{i \in \mathcal{F}}, \{c_{ij}\}_{i \in \mathcal{F}, j \in \mathcal{D}})$ of the LB knapsack median, pick a constant $\alpha \in [0, 1)$, get rid of the lower bounds $\{L_i\}_{i \in \mathcal{F}}$ from \mathcal{IN} and

$$\text{set } f_i := \frac{2\alpha}{1 - \alpha} \sum_{j \in \mathcal{D}_i} c_{ij} \text{ for every } i \in \mathcal{F},$$

in order to obtain instance $\mathcal{IN}' = (\mathcal{F}, \mathcal{D}, B, \{f_i\}_{i \in \mathcal{F}}, \{w_i\}_{i \in \mathcal{F}}, \{c_{ij}\}_{i \in \mathcal{F}, j \in \mathcal{D}})$ of the knapsack FL.

Step 2 Solve the instance of the knapsack FL.

Solve new instance \mathcal{IN}' with the current best η -approximation algorithm for the knapsack FL (see [4]), where $\eta = 17.46 + \epsilon$, and obtain a feasible solution (S', σ') .

Step 3 Construct a solution for the LB knapsack median.

Same as Step 3 in Algorithm 1. At the end of this step, output solution (S, σ) .

The following theorem gives the result for the LB knapsack median.

Theorem 3. *Algorithm 3 is a bi-criteria approximation algorithm for the LB knapsack median that produces a solution (S, σ) , which connects at least αL_i clients to every opened facility $i \in S$, and has an approximation ratio of $\frac{1+\alpha}{1-\alpha} \eta$ where $\alpha \in [0, 1)$ and $\eta = 17.46 + \epsilon$.*

We skip the proof of this theorem since it is analogous to the one for the LB k -median.

3.3 The Prize-Collecting Lower-Bounded k -median Problem

Compared with the LB k -median, in the PLB k -median, every client $j \in \mathcal{D}$ is given an additional penalty cost p_j . Our goal is to select at most k facilities to open, connect a portion of the clients and penalize the rest of them, so as to minimize the sum of opening, connection and penalty costs.

The algorithm for the PLB k -median is given by transforming all the steps in Algorithm 1.

Algorithm 4

Step 1 Construct a new instance of the P k -FL.

For the instance $\mathcal{IN} = (\mathcal{F}, \mathcal{D}, k, \{L_i\}_{i \in \mathcal{F}}, \{p_j\}_{j \in \mathcal{D}}, \{c_{ij}\}_{i \in \mathcal{F}, j \in \mathcal{D}})$ of the PLB k -median, pick a constant $\alpha \in [0, 1)$, get rid of the lower bounds $\{L_i\}_{i \in \mathcal{F}}$ from \mathcal{IN} and

$$\text{set } f_i := \frac{1 + \alpha}{1 - \alpha} \sum_{j \in \mathcal{D}_i} c_{ij} \text{ for every } i \in \mathcal{F},$$

in order to obtain instance $\mathcal{IN}' = (\mathcal{F}, \mathcal{D}, k, \{f_i\}_{i \in \mathcal{F}}, \{p_j\}_{j \in \mathcal{D}}, \{c_{ij}\}_{i \in \mathcal{F}, j \in \mathcal{D}})$ of the P k -FL.

Step 2 Solve the instance of the P k -FL.

Solve new instance \mathcal{IN}' with the current best θ -approximation algorithm for the P k -FL (see [18]), where $\theta = 2 + \sqrt{3} + \epsilon$, and obtain a feasible solution (S', P', σ') , where S' is the set of opened facilities, P' is the set of penalized clients and $\sigma' : \mathcal{D} \setminus P' \rightarrow S'$ is a function that maps every client $j \in \mathcal{D} \setminus P'$ to the closest facility in S' . For any client $j \in \mathcal{D} \setminus P'$, let $\sigma'(j)$ denote its closest facility in S' . For any client $j \in P'$, define its $\sigma'(j) := i_p$ where i_p is a dummy facility for penalizing.

Step 3 Construct a solution for the PLB k -median.

Step 3.1 Initialization.

At the very beginning, set $S := S'$, $P := P'$ and $\sigma(j) := \sigma'(j)$ for any $j \in \mathcal{D}$, define $T_i := \{j \in \mathcal{D} : \sigma(j) = i\}$, $l_i := |T_i|$ and $P_i := \{j \in \mathcal{D} : j \in \mathcal{D}_i, \sigma(j) = i_p\}$ for any $i \in \mathcal{F}$. Define $S_d := \{i \in S : l_i < \alpha L_i\}$.

Step 3.2 Close facilities and reconnect clients.

While $S_d \neq \emptyset$ **do**

Arbitrarily choose some facility $i \in S_d$. There are two possible cases.

Case 1. $|T_i| + |P_i| < \alpha L_i$.

In this case, close facility i . For every client $j \in T_i$, reconnect it to its closest facility $i' \in S \setminus \{i\}$ and update $\sigma(j) := i'$. Update $S := S \setminus \{i\}$. Update T_i, l_i for any facility $i \in \mathcal{F}$ and S_d .

Case 2. $|T_i| + |P_i| \geq \alpha L_i$.

In this case, for every client $j \in P_i$, connect it to its closest facility $i' \in S$ and update $\sigma(j) := i'$. Update $P := P \setminus P_i$. Then, update P_i, T_i as well as l_i for any facility $i \in \mathcal{F}$, also update S_d .

Output solution (S, P, σ) .

The following theorem provides the result for the PLB k -median.

Theorem 4. Algorithm 4 is a bi-criteria approximation algorithm for the PLB k -median that produces a solution (S, P, σ) , which connects at least αL_i clients to every opened facility $i \in S$, and has an approximation ratio of $\frac{2\theta}{1-\alpha}$ where $\alpha \in [0, 1)$ and $\theta = 2 + \sqrt{3} + \epsilon$.

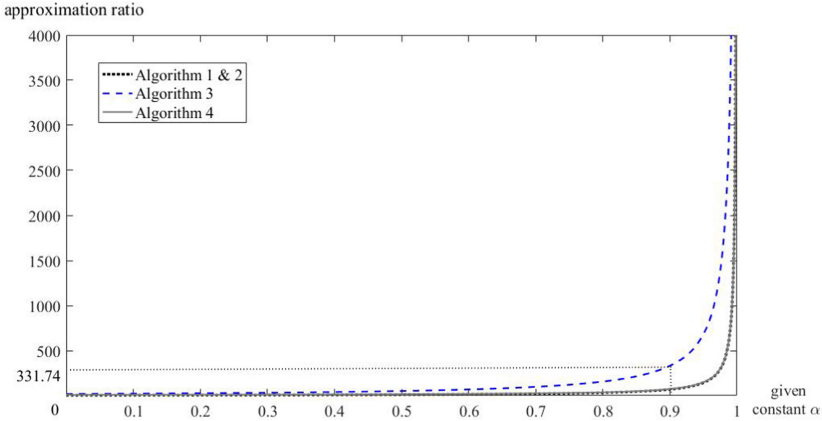


Fig. 1. Relationships between the constant α and approximation ratios.

The proof of this theorem is more intricate than the one for the LB k -median, since every client has an alternative decision in the PLB k -median, which is to be penalized. Suppose that (S^*, P^*, σ^*) is the optimal solution for the instance \mathcal{IN} of the PLB k -median. Let OPT_{plk} be the total cost of the solution (S^*, P^*, σ^*) for the instance \mathcal{IN} . From Algorithm 4, it is clear that the solution (S, P, σ) connects at least αL_i clients to every opened facility $i \in S$. We need the following lemmas to achieve the approximation ratio of our algorithm for the PLB k -median.

Lemma 4. *The total cost of the solution (S', P', σ') for the instance \mathcal{IN}' of the P k -FL is within a factor of $\frac{2\theta}{1-\alpha}$ of the total cost of the optimal solution (S^*, P^*, σ^*) for the instance \mathcal{IN} of the PLB k -median, i.e.,*

$$\sum_{i \in S'} f_i + \sum_{j \in \mathcal{D} \setminus P'} c_{\sigma'(j)j} + \sum_{j \in P'} p_j \leq \frac{2\theta}{1-\alpha} \cdot OPT_{plk},$$

where $\alpha \in [0, 1)$ and $\theta = 2 + \sqrt{3} + \epsilon$.

Lemma 5. *The total cost of the solution (S, P, σ) for the instance \mathcal{IN} of the PLB- k -median is no more than the total cost of the solution (S', P', σ') for the instance \mathcal{IN}' of the P k -FL, i.e.,*

$$\sum_{j \in \mathcal{D} \setminus P} c_{\sigma(j)j} + \sum_{j \in P} p_j \leq \sum_{i \in S'} f_i + \sum_{j \in \mathcal{D} \setminus P'} c_{\sigma'(j)j} + \sum_{j \in P'} p_j.$$

Combining Lemma 4 and Lemma 5 implies the approximation ratio of Algorithm 4.

4 Performances Evaluation of the Algorithms

In this section, through providing the relationships between the given constant $\alpha \in [0, 1)$ and the approximation ratios, we demonstrate the performances of Algorithm 1-4.

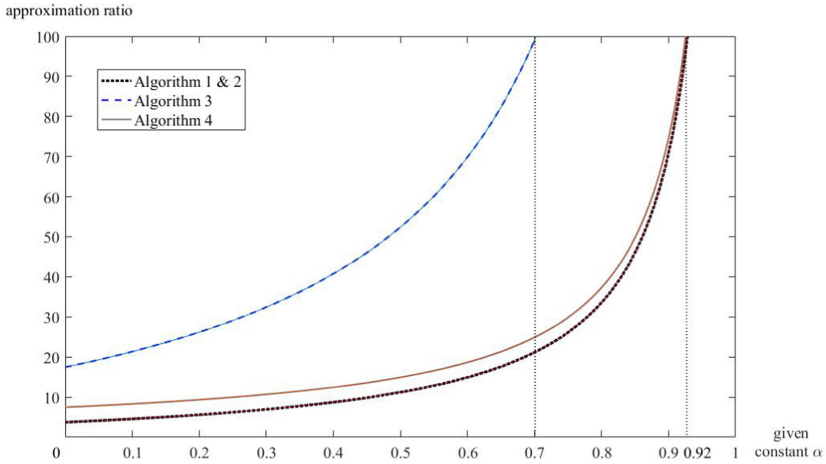


Fig. 2. Relationships between the constant α and approximation ratios. Zoom in on approximation ratios smaller than 100.

It is worth mentioning that, combining the idea behind Algorithm 1 and the significant first true 4000-approximation algorithm of Li [13] for the LBFL, we strongly believe that there exists $O(1)$ -approximation algorithm for the LB k -median which does not violate any lower bound. Unfortunately, the true approximation algorithm for the LB k -median cannot be practical, since the approximation ratio of it is likely no less than the one for the LBFL. Li [13] also states that even with more in-depth consideration, it is hard to reduce the approximation ratio for the LBFL to below 100 by using the same method.

Now, we want to demonstrate that our Algorithm 1 for the LB k -median can offer an obviously better approximation ratio while slightly violating the lower bound constraints. In Fig. 1, we show the relationships between constant $\alpha \in [0, 1)$ and the approximation ratios of Algorithm 1–4. Note that the approximation ratios grow slowly and steadily at first, but after the constant α exceeds 0.9 the approximation ratios begin to increase in a steep way. From Fig. 1, it is clear that when $\alpha = 0.9$ (i.e., when the algorithm outputs a solution that 90% of the lower bound requirement of the opened facility is satisfied), the approximation ratio of Algorithm 1 for the LB k -median is significantly better than 4000. Figure 2 zooms in on approximation ratios smaller than 100. As we can see, when Algorithm 1 outputs a solution for the LB k -median with an approximation ratio of 100, the solution can satisfy the majority (i.e., more than 90%) of the lower bound requirement of any opened facility. In some real-world applications, it would be advisable to choose our algorithm, which has a preferable approximation ratio and violates the lower bound constraints within an acceptable range (i.e., violates no more than 10 % of the lower bound requirements).

Additionally, Fig. 1 and Fig. 2 show that Algorithm 2 performs as same as Algorithm 1 and Algorithm 4 is almost as well as Algorithm 1. Among our algo-

rithms, Algorithm 3 has the worst performance. From Fig. 1, when $\alpha = 0.9$, the approximation ratio of Algorithm 3 for the LB knapsack median is visibly greater than the ratios of other algorithms. From Fig. 2, when Algorithm 3 outputs a solution with a ratio of 100, the solution can only guarantee to satisfy about 70% of the lower bound requirement of any opened facility.

Acknowledgement. The first author is supported by National Natural Science Foundation of China (No. 11531014). The second author is supported by National Natural Science Foundation of China (No. 11771003). The third author is supported by National Science Foundation of China (No. 11971349). The fourth author is supported by National Natural Science Foundation of China (No. 11871081) and the Science and Technology Program of Beijing Education Commission (No. KM201810005006).

References

1. Aggarwal, G., et al.: Achieving anonymity via clustering. *ACM Trans. Algorithms* **6**(3), 1–19 (2010)
2. Ahmadian, S., Swamy, C.: Improved approximation guarantees for lower-bounded facility location. In: Erlebach, T., Persiano, G. (eds.) *WAOA 2012*. LNCS, vol. 7846, pp. 257–271. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-38016-7_21
3. Arya, V., Garg, N., Khandekar, R., Meyerson, A., Munagala, K., Pandit, V.: Local search heuristics for k -median and facility location problems. *SIAM J. Comput.* **33**(3), 544–562 (2004)
4. Byrka, J., Pensyl, T., Rybicki, B., Spoerhase, J., Srinivasan, A., Trinh, K.: An improved approximation algorithm for knapsack median using sparsification. *Algorithmica* **80**(4), 1093–1114 (2018). <https://doi.org/10.1007/s00453-017-0294-4>
5. Byrka, J., Pensyl, T., Rybicki, B., Srinivasan, A., Trinh, K.: An improved approximation for k -median, and positive correlation in budgeted optimization. *ACM Trans. Algorithms* **13**(2), 1–31 (2017)
6. Charikar, M., Guha, S., Tardos, É., Shmoys, D.B.: A constant-factor approximation algorithm for the k -median problem. *J. Comput. Syst. Sci.* **65**(1), 129–149 (2002)
7. Guha, S., Meyerson, A., Munagala, K.: Hierarchical placement and network design problems. In: *Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, pp. 603–612 (2000)
8. Jain, K., Mahdian, M., Markakis, E., Saberi, E., Vazirani, V.V.: Greedy facility location algorithms analyzed using dual fitting with factor-revealing LP. *J. ACM* **50**(6), 795–824 (2003)
9. Jain, K., Mahdian, M., Saberi, A.: A new greedy approach for facility location problems. In: *Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, pp. 731–740 (2002)
10. Jain, K., Vazirani, V.V.: Approximation algorithms for metric facility location and k -median problems using the primal-dual schema and Lagrangian relaxation. *J. ACM* **48**(2), 274–296 (2001)
11. Karger, D.R., Minkoff, M.: Building Steiner trees with incomplete global knowledge. In: *Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, pp. 613–623 (2000)
12. Li, S.: A 1.488 approximation algorithm for the uncapacitated facility location problem. *Inf. Comput.* **222**, 45–58 (2013)

13. Li, S.: On facility location with general lower bounds. In: Proceedings of the 30th Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 2279–2290 (2019)
14. Li, S., Svensson, O.: Approximating k -median via pseudo-approximation. *SIAM J. Comput.* **45**(2), 530–547 (2016)
15. Shmoys, D.B., Tardos, É., Aardal, K.I.: Approximation algorithms for facility location problems. In: Proceedings of the 29th Annual ACM symposium on Theory of Computing, pp. 265–274 (1997)
16. Sviridenko, M.: An improved approximation algorithm for the metric uncapacitated facility location problem. In: Cook, W.J., Schulz, A.S. (eds.) IPCO 2002. LNCS, vol. 2337, pp. 240–257. Springer, Heidelberg (2002). https://doi.org/10.1007/3-540-47867-1_18
17. Svitkina, Z.: Lower-bounded facility location. *ACM Trans. Algorithms* **6**(4) (2010). Article no. 69
18. Wang, Y., Xu, D., Du, D., Wu, C.: An approximation algorithm for k -facility location problem with linear penalties using local search scheme. *J. Comb. Optim.* **36**(1), 264–279 (2016). <https://doi.org/10.1007/s10878-016-0080-2>
19. Zhang, P.: A new approximation algorithm for the k -facility location problem. *Theoret. Comput. Sci.* **384**(1), 126–135 (2007)