An Improved Algorithm for the Hard Uniform Capacitated k-median Problem

Shanfei Li

Delft Institute of Applied Mathematics, TU Delft, The Netherlands
shanfei.li@tudelft.nl

Abstract

In the k-median problem, given a set of locations, the goal is to select a subset of at most k centers so as to minimize the total cost of connecting each location to its nearest center. We study the uniform hard capacitated version of the k-median problem, in which each selected center can only serve a limited number of locations.

Inspired by the algorithm of Charikar, Guha, Tardos and Shmoys, we give a $(6 + 10\alpha)$-approximation algorithm for this problem with increasing the capacities by a factor of $2+\frac{2}{\alpha}, \alpha \geq 4$, which improves the previous best $(32l^2 + 28l + 7)$-approximation algorithm proposed by Byrka, Fleszar, Rybicki and Spoerhase violating the capacities by factor $2 + \frac{3}{l-1}, l \in \{2, 3, 4, \dots\}$.

1998 ACM Subject Classification F.2.2 Nonnumerical Algorithms and Problems

Keywords and phrases Approximation algorithm, k-median problem, LP-rounding, Hard capacities

Digital Object Identifier 10.4230/LIPIcs.APPROX-RANDOM.2014.325

1 Introduction

In the capacitated k-median problem (CKM), we are given a set N of locations (where a center can potentially be opened). Each location $j \in N$ has a capacity M (uniform capacities), and a demand d_j that must be served. Assigning one unit of the demand of location j to center $i \in N$ incurs service costs c_{ij}. We assume the service costs are non-negative, identity of indiscernibles, symmetric and satisfy the triangle inequality. That is, $c_{ij} \geq 0, \forall i, j \in N$; $c_{ij} = 0$, if $i = j$; $c_{ij} = c_{ji}, \forall i, j \in N$ and $c_{it} + c_{tj} \geq c_{ij}, \forall i, j, t \in N$. The objective is to serve all the demands by opening at most k centers and satisfying the capacity constraints such that the total cost is minimized. In this paper, we consider the *hard* capacities and *splittable* demands, that is, we allow at most one center to be opened at any location and each location can be served from more than one open center. (In contrast, the *soft* capacities allows that multiple centers can be opened in a single location. In the *unsplittable* demands case each location must be served by exactly one open center.)

CKM can be formulated as the following mixed integer program (MIP), where variable x_{ij} indicates the fraction of the demand of location j that is served by location i, and y_i indicates whether location i is selected as a center.

$$\begin{aligned}
\min \quad & \sum_{i,j \in N} d_j c_{ij} x_{ij} \\
\text{subject to:} \quad & \sum_{i \in N} x_{ij} = 1, \quad \forall j \in N; \quad \sum_{j \in N} d_j x_{ij} \leq M y_i, \quad \forall i \in N; \\
& \sum_{i \in N} y_i \leq k; \qquad\qquad\quad 0 \leq x_{ij} \leq y_i, \quad \forall i, j \in N; \\
& y_i \in \{0, 1\}, \quad \forall i \in N. \qquad\qquad\qquad\qquad\qquad\qquad (1)
\end{aligned}$$

Replacing constraints (1) by $0 \leq y_i \leq 1, \forall i \in N$, we obtain the LP-relaxation of CKM.

17th Int'l Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX'14) / 18th Int'l Workshop on Randomization and Computation (RANDOM'14).
Editors: Klaus Jansen, José Rolim, Nikhil Devanur, and Cristopher Moore; pp. 325–338
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1.1 Related Work and Our Results

The k-median problem is a classical NP-hard problem in computer science and operations research, and has a wide variety of applications in clustering and data mining [4, 13]. The uncapacitated k-median problem was studied extensively [1, 2, 6, 8, 9, 14, 15, 17], and the best known approximation algorithm was recently given by Byrka et al. [6] with approximation ratio $2.611 + \epsilon$ by improving the algorithm of Li and Svensson [17].

The capacitated versions of k-median problem are much less understood. The above LP-relaxation has an unbounded integrality gap. More precisely, the capacity or the number of opened centers has to be increased by a factor of at least 2, if we try to get an integral solution within a constant factor of the cost of an optimal solution to the LP-relaxation [9]. All the previous attempts with constant approximation ratios for this problem violate at least one of the two kinds of hard constraints: the capacity constraint and cardinality constraint (at most k centers can be opened), even the local search technique.

For the hard uniform capacity case, by increasing the capacities within a factor of 3, Charikar et al. [7, 9, 12] presented a 16-approximation algorithm based on LP-rounding. This violation ratio of capacities was recently improved to $2 + \frac{3}{l-1}$, $l \in \{2, 3, 4, \dots\}$ by Byrka et al. [5], with the corresponding approximation ratio of $32l^2 + 28l + 7$. In addition, Korupolu et al. [16] proposed a $(1 + 5/\epsilon)$-approximation algorithm while opening at most $(5 + \epsilon)k$ centers, and a $(1 + \epsilon)$-approximation algorithm while opening at most $(5 + 5/\epsilon)k$ centers based on a local search technique.

For soft non-uniform capacities, Chuzhoy and Rabani [10] presented a 40-approximation algorithm while violating the capacities within a factor of 50 based on primal-dual and Lagrangian relaxation methods. Using at most $(1 + \delta)k$ facilities, Bartal et al. [3] gave a $19.3(1 + \delta)/\delta^2$-approximation algorithm ($\delta > 0$). For hard non-uniform capacities, Gijswijt and Li [11] gave a $(7 + \epsilon)$-approximation algorithm while opening at most $2k$ centers.

In this paper, we improve the algorithm of Charikar et al. [9] to reduce its violation ratio of capacities from 3 to $2+\frac{2}{\alpha}, \alpha \geq 4$ and get an $(6+10\alpha)$-approximation algorithm for the hard uniform capacitated k-median problem, which improves the previous best approximation ratio for any violation ratio of capacities in $(2, 3)$. The approximation ratios we obtain for violation ratio of 2.1, 2.3, 2.5, 2.75 and 3 (for instance) are summarized in the following table.

violation ratio of capacities	2.1	2.3	2.5	2.75	3
previous best	31627	4187	1771	947	16
our algorithm	206	72.67	46	46	46

Note that with increasing the capacities by a factor of at least 3, the best approximation ration is still due to Charikar et al. [9].

Additionally, for metric facility location problems there is a slightly different model for the capacitated k-median [5, 11], in which we are given a set F of facilities and a set D of clients. Each facility has a capacity M. Each client $j \in D$ has a demand d_j that has to be served by facilities. Note that the capacity of each client is 0. This is different from our model, in which each location has a capacity M. We show that our algorithm can be easily extended to solve this model with increasing the approximation ratio by a factor at most $2 + \frac{1}{6+10\alpha}$.

1.2 The Main Idea Behind Our Algorithm

In Charikar et al. [9] algorithm, based on an optimal solution to the LP-relaxation, a $\{\frac{1}{2}, 1\}$-solution (x, y) is first constructed such that $y_i \in \{0, \frac{1}{2}, 1\}, \forall i \in N$; $\sum_{j \in N} x_{ij} d_j \leq M$,

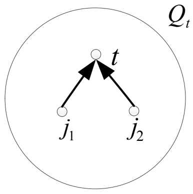

Figure 1 *A star Q_t.*

if $y_i = \frac{1}{2}$; and $\sum_{j \in N} x_{ij} d_j \le 2M$, if $y_i = 1$. Note that $\sum_{j \in N} x_{ij} d_j \le M y_i$ could be violated in this solution.

Next, a center is directly opened at location i if $y_i = 1$. Then, they construct a collection of rooted stars spanning the locations $i \in N$ with $y_i = \frac{1}{2}$. By a star by star rounding procedure, exactly half of the locations with fractional opening value $\frac{1}{2}$ are chosen as centers. The demands of another half of the locations, where no center is opened finally, are reassigned to the opened half. In the worst case, the capacity of the root of some star has to be increased by factor 3 to satisfy the capacity constraint. Take Fig. 1 as an example. The star Q_t, rooted at t, has two children j_1 and j_2 with $y_t = y_{j_1} = y_{j_2} = \frac{1}{2}$. In the worst case of Charikar et al. algorithm, we are allowed to build at most $\lfloor y_t + y_{j_1} + y_{j_2} \rfloor$ centers, i. e., 1 center. Without loss of generality, suppose we build a center at the root t, and reassign the demand served by j_1 and j_2 to t. Then, the capacity of t has to be increased by factor 3 to satisfy the capacity constraint, as $\sum_{j \in N} x_{ij} d_j \le M$ for $i = t, j_1, j_2$.

We generalize the algorithm of Charikar et al. to improve its violation ration from 3 to $2 + \epsilon$. The key idea behind our algorithm relies on the following observations. One is that if we can obtain a $\{1 - \frac{1}{\delta}, 1\}$-solution, then 2 centers can be built for the above example in the worst case by setting $\delta \ge 3$, as then $\lfloor y_t + y_{j_1} + y_{j_2} \rfloor \ge \lfloor \frac{2}{3} + \frac{2}{3} + \frac{2}{3} \rfloor = 2$. Consequently, we only need to blow up the capacity of location t by factor 2 instead of 3, by building centers at t and j_2, and assigning the demand served by j_1 to t. However, this example only shows one kind of stars. To make sure the violation ratio can be improved for all kinds of stars, we construct a $\{(\frac{\alpha-2}{\alpha}, \frac{\alpha-1}{\alpha}], [1, 2)\}$-solution (x, y) such that

1. for each $i \in N$, $\frac{\alpha-2}{\alpha} < y_i \le \frac{\alpha-1}{\alpha}$, or $1 \le y_i < 2$, or $y_i = 0$; and $|\{i \in N \mid \frac{\alpha-2}{\alpha} < y_i < \frac{\alpha-1}{\alpha}\}| \le 1$;
2. if $\frac{\alpha-2}{\alpha} < y_i \le \frac{\alpha-1}{\alpha}$, then $\sum_{j \in N} d_j x_{ij} \le M$;
3. if $1 \le y_i < 2$, then $\sum_{j \in N} d_j x_{ij} \le M y_i$.

Another one is that constraints $y_i \le 1, \forall i \in N$ hold in each step of the algorithm by Charikar et al. That is, they round $y_i > 1$ to be 1 for each $i \in N$ in each step. This is a quite natural operation since we consider the hard capacitated case, i. e., at most one center can be opened at any location. However, we observe that after obtaining an optimal solution to the LP-relaxation, it is sufficient to make sure constraints $y_i \le 1, \forall i \in N$ hold in our last step. For all other steps (except last step), this rounding can be avoided by relaxing the constraint $y_i \le 1$ to $y_i < 2$. We use an example to show the profit we can gain from avoiding this rounding. Suppose we have a star Q_t rooted at t with one child j_1. Moreover, $y_t = 1.9$ and $y_{j_1} = 0.5$. Then, in the worst case, we can build $\lfloor y_t + y_{j_1} \rfloor = 2$ centers. We open t and j_1. Consequently, we only need to increase the capacity of t by factor 1.9 (note that if $1 \le y_i < 2$, then $\sum_{j \in N} d_j x_{ij} \le M y_i$ for our $\{(\frac{\alpha-2}{\alpha}, \frac{\alpha-1}{\alpha}], [1, 2)\}$-solution). However, if we round 1.9 to 1, we obtain a star Q_t with $y_t = 1$ and $y_{j_1} = 0.5$. Then, in the worst case, only 1 center can be built as $\lfloor y_t + y_{j_1} \rfloor = 1$. Without loss of generality, suppose we build a center at t, and assign the demand served by j_1 to t. Then, we need to increase the capacity of t by factor 2.9.

2 An Improved Approximation Algorithm

From now on, let (x, y) denote an optimal solution to the LP-relaxation with total cost C_{LP}. We consider y_i as the *opening value* of location i. If $y_i \in (0,1)$, we say that location i is *fractionally opened* (as a center). For each $j \in N$, define $C_j = \sum_{i\in N} c_{ij}x_{ij}$. Note that $C_{LP} = \sum_{j\in N} d_j C_j$. The outline of our algorithm is similar to [9].

Step 1. We partition locations to a collection of clusters. The total opening value of each cluster is at least $\frac{\alpha-1}{\alpha}, \alpha \geq 4$.

Step 2. For each cluster, we integrate the nearby opened locations to obtain a $[\frac{\alpha-1}{\alpha}, 2)$-solution (x', y') to the LP-relaxation, which satisfies the relaxing constraints $0 \leq y'_i < 2$ instead of $0 \leq y'_i \leq 1$ for each $i \in N$.

Step 3. We redistribute the opening values among locations with $y'_i \in [\frac{\alpha-1}{\alpha}, 1)$ to obtain a $\{(\frac{\alpha-2}{\alpha}, \frac{\alpha-1}{\alpha}], [1,2)\}$-solution $(x', \hat{y})$, which satisfies the relaxing constraints $\sum_{j\in N} d_j x'_{ij} \leq M$ if $\hat{y}_i \in (0,1)$, $\sum_{j\in N} d_j x'_{ij} \leq M\hat{y}_i$ otherwise, instead of $\sum_{j\in N} d_j x'_{ij} \leq M\hat{y}_i$ for each $i \in N$.

Step 4. We round the $\{(\frac{\alpha-2}{\alpha}, \frac{\alpha-1}{\alpha}], [1,2)\}$-solution to be an integral solution with increasing the capacities by a factor of $2 + \frac{2}{\alpha}$.

2.1 Step 1: Clustering

In this step, by the *filtering* technique of Lin and Vitter [18], we will partition locations into clusters, and for each cluster select a single location as the *core* of this cluster, such that each location in the cluster is not far to its cluster core and the cores are sufficiently far to each other.

Let N' be the collection of all cluster cores. Let $N'(j)$ denote the closest cluster core to j in N'. For each $l \in N'$, let M_l denote the cluster whose core is l, and define $Z_l = \sum_{j\in M_l} y_j$ be the total opening value of all locations in cluster M_l.

▶ **Definition 1.** We call a cluster M_l *terminal* if $Z_l \geq 1$, *non-terminal* if $0 < Z_l < 1$.

Let $n = |N|$. The clustering is done by Procedure 1 (similar to [9]). After this step, the following properties hold ($\alpha \geq 4$):

[**1a**]. $\forall j \in M_l, l \in N'$, $c_{lj} \leq 2\alpha C_j$;
[**1b**]. $\forall l, l' \in N'$ and $l \neq l'$, $c_{ll'} > 2\alpha \max\{C_l, C_{l'}\}$;
[**1c**]. $\forall l \in N', Z_l = \sum_{j\in M_l} y_j \geq \frac{\alpha-1}{\alpha}$;
[**1d**]. $\bigcup_{l\in N'} M_l = N$; and $M_l \bigcap M_{l'} = \emptyset, \forall l, l' \in N'$ and $l \neq l'$.

We can easily get property **1a**, **1b** and **1d** from this procedure.

Note that location i belongs to cluster M_l if $c_{il} \leq \alpha C_l$. For contradiction, suppose for some $i \in N$ with $c_{il} \leq \alpha C_l$, $i \in M_{l'}$ instead of $i \in M_l$, where $l' \in N' - \{l\}$. This means $c_{il'} \leq c_{il}$ as we add i to cluster $M_{l'}$ only if $N'(i) = l'$. Then, we have $c_{ll'} \leq c_{il} + c_{il'} \leq 2c_{il} \leq 2\alpha C_l$, which is a contradiction as $c_{ll'} > 2\alpha C_l$ by property **1b**. Then, we have the following lemma. See [18] for the proof.

▶ **Lemma 2.** (property **1c**) $\forall l \in N', Z_l \geq \frac{\alpha-1}{\alpha}$.

2.2 Step 2: Obtaining a $[\frac{\alpha-1}{\alpha}, 2)$-solution

We will get rid of locations with relatively small fractional opening value in this step, by constructing a $[\frac{\alpha-1}{\alpha}, 2)$-solution (x', y') in which $y'_i = 0$ or $\frac{\alpha-1}{\alpha} \leq y'_i < 2$, $\forall i \in N$. For each cluster M_l, we transfer the amount of locations (their opening values and the demands served by these locations) far away from the cluster core l to locations closer to l.

Procedure 1. Clustering

1. order all locations in nondecreasing order of C_j, (without loss of generality, assume $C_1 \leq \cdots \leq C_n$);
2. set $N' := \emptyset$;
3. **for** $j = 1$ *to* n **do**
    find a location $l \in N'$ such that $c_{lj} \leq 2\alpha C_j$, where $\alpha \geq 4$;
    if *no such location is found* **then**
        choose j as a cluster core, i. e., set $N' := N' \cup \{j\}$;
    end
end
4. set $M_l := \emptyset$, $\forall l \in N'$;
5. **for** $j = 1$ *to* n **do**
    if *j is closer to cluster core $l \in N'$ than all other cluster cores (break ties arbitrarily)* **then**
        add location j to cluster M_l. (i. e., set $M_l := \{j \in N \mid N'(j) = l\}$.)
    end
end

In this step, initially set $y'_i = y_i, x'_{ij} = x_{ij}, \forall i, j \in N$. Then, we consider clusters one by one. For each cluster $M_l, l \in N'$, order locations in M_l in nondecreasing value of $c_{lj}, j \in M_l$. Without loss of generality, assume we get an order $j_1, \cdots, j_u$ (note that $j_1 = l$). If we decide to move the amount of location j_b to j_a $(1 \leq a < b \leq u)$, then perform Procedure 2 [7, 12].

Procedure 2. Move(j_a,j_b)

1. let $\delta = \min\{1 - y'_{j_a}, y'_{j_b}\}$;
2. for all $j \in N$, set $x'_{j_a j} := x'_{j_a j} + \frac{\delta}{y'_{j_b}} x'_{j_b j}$, $x'_{j_b j} := x'_{j_b j} - \frac{\delta}{y'_{j_b}} x'_{j_b j}$;
3. set $y'_{j_a} := y'_{j_a} + \delta$,$y'_{j_b} := y'_{j_b} - \delta$;

▶ **Lemma 3.** *After Procedure 2, we still have*
1. $\sum_{j \in M_l} y'_j = \sum_{j \in M_l} y_j$, *for each* $l \in N'$;
2. *for each* $j \in N$, $\sum_{i \in N} x'_{ij} = 1$;
3. $\sum_{j \in N} d_j x'_{ij} \leq M y'_i$, *for each* $i \in N$.

We use Procedure 3 to decide whether we move the amount of location j_b to j_a.

▶ **Lemma 4.** *If in Procedure 3, j_a exists but j_b does not exist, and M_l is a terminal cluster, then $a \geq 2$ and $y'_{j_{a-1}} = 1$.*

Proof. Since M_l is a terminal cluster, we have $Z_l \geq 1$. Moreover, we know $y'_{j_t} = 1$ for each $t < a$ and $y'_{j_s} = 0$ for each $s > a$, as j_b does not exist. Thus, $a \geq 2$. Otherwise, $Z_l < 1$, a contradiction. ◀

▶ **Lemma 5.** *After this step, we have the following properties*
[**2a**]. *for all $i \in N$, $\frac{\alpha-1}{\alpha} \leq y'_i < 2$ or $y'_i = 0$; and $\sum_{j \in N} d_j x'_{ij} \leq M y'_i$;*
[**2b**]. $\sum_{i \in N} y'_i = \sum_{i \in N} y_i \leq k$;
[**2c**]. $x'_{ij} \leq y'_i, \forall i, j \in N$.

Procedure 3. Concentrate(M_l)

while *there exists a location in M_l with fractional opening value* **do**
  1. let j_a be the first location in the sequence $j_1, \cdots, j_u$ such that $0 \leq y'_{j_a} < 1$;
  2. let j_b be the first location in the sequence $j_{a+1}, \cdots, j_u$ such that $0 < y'_{j_b} \leq 1$;
  3. **if** j_a *and* j_b *both exist* **then**
    execute procedure Move(j_a,j_b) to move the amount of j_b to j_a;
  end
  4. **if** j_a *exists but* j_b *does not exist* **then**
    if M_l *is a terminal cluster,i. e.,*$a \geq 2$ **then**
      set $y'_{j_{a-1}} := y'_{j_{a-1}} + y'_{j_a}, y'_{j_a} := 0$;
      for each $j \in N$, set $x'_{j_{a-1}j} := x'_{j_{a-1}j} + x'_{j_a j}, x'_{j_a j} := 0$;
    end
    terminate.
  end
end

Proof. Property **2a.** If M_l is a non-terminal cluster, i. e., $0 < Z_l < 1$, then we will move the amount of each location in M_l to its core l according to Procedure 3. Consequently, we obtain $\frac{\alpha-1}{\alpha} \leq y'_l = Z_l < 1$ (property **1c**) and $y'_j = 0, \forall j \in M_l - \{l\}$.

If M_l is a terminal cluster, i. e., $Z_l \geq 1$, then according to Lemma 4 we get $y'_{j_t} = 1$ for each $t < a$ and $y'_{j_s} = 0$ for each $s > a$ if j_a exists and j_b does not exist. Then, we move the amount of y'_{j_a} to $y'_{j_{a-1}}$. So, $1 \leq y'_{j_{a-1}} < 2$ as $0 \leq y'_{j_a} < 1$. Note that if j_a does not exist, we know $y'_j = 1$ for each $j \in M_l$.

Thus, for all $i \in N$, $\frac{\alpha-1}{\alpha} \leq y'_i < 2$ or $y'_i = 0$. $\sum_{j \in N} d_j x'_{ij} \leq M y'_i, \forall i \in N$ hold by Lemma 3 (note that it is easy to check these inequalities still hold after the step 4 in Procedure 3).

Property **2b.** This directly follows by Lemma 3(1).

Property **2c.** Initially, we set $y'_i = y_i, x'_{ij} = x_{ij}$ for all $i, j \in N$. Thus, $x'_{ij} \leq y'_i$ holds, for each $i, j \in N$. We will show that after the procedure these inequalities still hold.

For each non-terminal cluster, only the core has a positive opening value after this step. And in the procedure the opening value of core is always increased by a bigger amount than the increasing of the fraction of the demand served by the core.

For a terminal cluster, each location i in the cluster has $y'_i = 0$ or $y'_i \geq 1$ after this step. Note that for each location $i \in N$ with $y'_i \geq 1$, $x'_{ij} \leq y'_i$ holds for each $j \in N$ as $x'_{ij} \leq 1$. Moreover, observe that for each $j \in N$, we always set $x'_{ij} := 0$ if y'_i is already set to be 0. ◀

Since each location is not far away from its cluster core, these transfer operations would not increase too much extra cost. More precisely, we can bound the service cost by the following lemma. The proof is similar as Lemma 2.8.3 and 2.8.3 in [7].

▶ **Lemma 6.** *(1). Let M_l be a non-terminal cluster. The demand of location j originally served by $j_b (j_b \in M_l)$ must be served by core l after the procedure. And we have $c_{lj} \leq 2c_{j_b j} + 2\alpha C_j$.*

(2). Let M_l be a terminal cluster. If we move the demand of location j served by j_b to j_a ($j_a, j_b \in M_l$, $a < b$), we have $c_{j_a j} \leq 3c_{j_b j} + 4\alpha C_j$.

Let $N_1 = \{i \in N \mid y'_i \geq 1\}$ be the collection of locations with the opening value at least 1. Let $N_2 = \{i \in N \mid y'_i \in [\frac{\alpha-1}{\alpha}, 1)\}$ be the collection of locations with fractional opening value in $[\frac{\alpha-1}{\alpha}, 1)$. Note that N_2 can also be written as $\{i \in N' \mid Z_i \in [\frac{\alpha-1}{\alpha}, 1)\}$. That is, N_2 is the collection of non-terminal cluster cores. Moreover, we have $N_1 \cup N_2 \supseteq N'$.

▶ **Lemma 7.** *If $|N_2| - 1 < \sum_{i\in N_2} y'_i$, we can get an integer solution with increasing the capacity by factor 2, by opening all locations in $N_1 \cup N_2$ as centers. The total cost of the obtained solution can be bounded by $(3 + 4\alpha)C_{LP}$.*

Proof. If $|N_2|-1 < \sum_{i\in N_2} y'_i$, then $|N_2| = \lceil \sum_{i\in N_2} y'_i \rceil$ as $y'_i < 1$ for each $i \in N_2$. Additionally, since $\sum_{i\in N_1} y'_i \le k - \sum_{i\in N_2} y'_i$ (by property **2b**) and $y'_i \ge 1$ for each $i \in N_1$, we have $|N_1| \le \lfloor k - \sum_{i\in N_2} y'_i \rfloor$.

Thus, if we only open locations in $N_1 \cup N_2$, then we open at most k centers as $\lceil \sum_{i\in N_2} y'_i \rceil + \lfloor k - \sum_{i\in N_2} y'_i \rfloor = k$.

Since $y'_i = 0$ for each $i \notin N_1 \cup N_2$, we have $\sum_{i\in N_1\cup N_2} x'_{ij} = 1, \forall j \in N$ by Lemma 3(2) and property **2c.** That is, $\sum_{i\in N_1\cup N_2} d_j x'_{ij} = d_j$ for each $j \in N$. Thus, the demand of each $j \in N$ can be satisfied by assigning $d_j x'_{ij}$ to $i \in N_1 \cup N_2$.

By Lemma 6, it is easy to see that the total cost of the obtained solution can be bounded by $(3 + 4\alpha)C_{LP}$. By Lemma 5, we know for all $i \in N$, $\frac{\alpha-1}{\alpha} \le y'_i < 2$ or $y'_i = 0$; and $\sum_{j\in N} d_j x'_{ij} \le M y'_i$. So, we increase the capacity by at most a factor of 2. ◀

From now on, we only consider the following case.

▶ **Assumption 8.** $\sum_{i\in N_2} y'_i \le |N_2| - 1$.

▶ **Definition 9.** We define new demands d' as follows. For each $i \in N$, set $d'_i := \sum_{j\in N} d_j x'_{ij}$. (Note that $d'_i = 0$ for each $i \in N - (N_1 \cup N_2)$.)

2.3 Step 3: Obtaining a $\{(\frac{\alpha-2}{\alpha}, \frac{\alpha-1}{\alpha}], [1, 2)\}$-solution

For each $i \in N_2$, let $s(i)$ be the nearest location to i in $(N_1 \cup N_2) - \{i\}$ (break ties arbitrarily). Let $Y = \sum_{i\in N_2} y'_i$. Note that we only consider the case: $Y \le |N_2| - 1$ by Assumption 8. After this step we will obtain a solution $(x', \hat{y})$ with $\frac{\alpha-2}{\alpha} < \hat{y}_i \le \frac{\alpha-1}{\alpha}$, or $1 \le \hat{y}_i < 2$, or $\hat{y}_i = 0$ for each $i \in N$.

In this step, initially we order all locations in N_2 in nondecreasing order of $d'_i c_{s(i)i}$. Without loss of generality, suppose we get an order $i_1, \cdots, i_v$. Next, for each $i \in N - N_2$, set $\hat{y}_i := y'_i$. For each $i \in N_2$, set $\hat{y}_i := \frac{\alpha-1}{\alpha}$. Let $Y' := Y - \sum_{i\in N_2} \hat{y}_i$. Then, perform Procedure 4.

▶ Remark. The Procedure 4 terminates at $r > 1$. If the procedure terminates at $r = 1$, then we get $Y = \sum_{t=1}^{v} y'_{i_t} > |N_2| - 1$, a contradiction.

▶ **Lemma 10.** *After the above procedure, we have the following properties*

[**3a**]. *for all $i \in N$, $\frac{\alpha-2}{\alpha} < \hat{y}_i \le \frac{\alpha-1}{\alpha}$, or $1 \le \hat{y}_i < 2$, or $\hat{y}_i = 0$; and only $\hat{y}_{i_1}$ can be in $(\frac{\alpha-2}{\alpha}, \frac{\alpha-1}{\alpha})$, i. e., $|\{i \in N \mid \frac{\alpha-2}{\alpha} < \hat{y}_i < \frac{\alpha-1}{\alpha}\}| \le 1$;*

[**3b**]. *for any location $i \in N$, if $\frac{\alpha-2}{\alpha} < \hat{y}_i \le \frac{\alpha-1}{\alpha}$, then $d'_i = \sum_{j\in N} d_j x'_{ij} \le M$;*

[**3c**]. *for any location $i \in N$, if $1 \le \hat{y}_i < 2$, then $d'_i = \sum_{j\in N} d_j x'_{ij} \le M\hat{y}_i$;*

[**3d**]. *$\sum_{i\in N_2} \hat{y}_i = \sum_{i\in N_2} y'_i$; $\sum_{i\in N} \hat{y}_i = \sum_{i\in N} y'_i \le k$;*

[**3e**]. *$\sum_{i\in N_2}(1 - \hat{y}_i) d'_i c_{s(i)i} \le \sum_{i\in N_2}(1 - y'_i) d'_i c_{s(i)i}$.*

Proof. Property **3a**. For each location $i \in N - N_2$, we set $\hat{y}_i := y'_i$. So, $1 \le \hat{y}_i < 2$ for each $i \in N_1$; $\hat{y}_i = 0$ for each $i \in N - (N_1 \cup N_2)$.

For each location $i \in N_2$, initially we set $\hat{y}_i := \frac{\alpha-1}{\alpha}$. In the Procedure 4, only $\hat{y}_{i_1}$ could be decreased by a number in $(0, \frac{1}{\alpha})$. The opening value of other location in N_2 remains the same or is set to be 1.

Procedure 4. Determine new opening values for $N_2(Y \leq |N_2| - 1)$

```
for r = v to 1 do
    if Y' = 0 then
        terminate;
    end
    if Y' > 0 and Y' + ŷ_{i_r} < 1 then
        set ŷ_{i_1} := ŷ_{i_1} − (1 − Y' − ŷ_{i_r}), ŷ_{i_r} := 1;
        terminate;
    end
    if Y' > 0 and Y' + ŷ_{i_r} ≥ 1 then
        set ŷ_{i_r} := 1 and update Y' := Y − Σ_{i∈N_2} ŷ_i;
    end
end
```

Property **3b, 3C**. Notice that if for location i we have $\frac{\alpha-2}{\alpha} < \hat{y}_i \leq \frac{\alpha-1}{\alpha}$ after the procedure, then we know $\frac{\alpha-1}{\alpha} \leq y'_i < 1$. And if $1 \leq \hat{y}_i < 2$ for location i after the procedure, then we have $y'_i \leq \hat{y}_i$.

We make no change on x'. Thus, combining with property **2a**, we have if $\frac{\alpha-2}{\alpha} < \hat{y}_i \leq \frac{\alpha-1}{\alpha}$, then $\sum_{j\in N} d_j x'_{ij} \leq M y'_i < M$. If $1 \leq \hat{y}_i < 2$, then $\sum_{j\in N} d_j x'_{ij} \leq M y'_i \leq M\hat{y}_i$.

Property **3d**. We move the opening value from one location to the other locations. We do not change the total opening value. So, $\sum_{i\in N_2} \hat{y}_i = \sum_{i\in N_2} y'_i$ holds after Procedure 4. Moreover, we set $\hat{y}_i := y'_i$ for each $i \in N - N_2$. Thus, we also have $\sum_{i\in N} \hat{y}_i = \sum_{i\in N} y'_i \leq k$.

Property **3e**. We always transfer the opening value from i_a to i_b, where $a < b$ and $d'_{i_b} c_{s(i_b)i_b} \geq d'_{i_a} c_{s(i_a)i_a}$. Therefore, $\sum_{i\in N_2} \hat{y}_i d'_i c_{s(i)i} \geq \sum_{i\in N_2} y'_i d'_i c_{s(i)i}$. Then, we have $\sum_{i\in N_2}(1-\hat{y}_i) d'_i c_{s(i)i} \leq \sum_{i\in N_2}(1-y'_i) d'_i c_{s(i)i}$. ◀

2.4 Step 4: Rounding to an Integral Solution

Let $\hat{N}_1 = \{i \in N \mid 2 > \hat{y}_i \geq 1\}$ be the set of locations with opening value greater than or equal to 1. Let $\hat{N}_2 = \{i \in N \mid \frac{\alpha-2}{\alpha} < \hat{y}_i \leq \frac{\alpha-1}{\alpha}\}$ be the set of location with fractional opening value strictly less than 1. Let $L_1 = |\hat{N}_1|$. Note that $N_1 \cup N_2 = \hat{N}_1 \cup \hat{N}_2$, and $\hat{N}_2 \subseteq N_2$.

In this step, we aim to construct an integral solution $(\bar{x}, \bar{y})$ with $\sum_{j\in N} \bar{x}_{ij} d'_j \leq (2+\frac{2}{\alpha}) M \bar{y}_i$ for each $i \in N$. If location j is opened as a center, we serve the demand d'_j of location j by itself. That is, set $\bar{x}_{jj} := 1, \bar{x}_{ij} := 0$ for each $i \neq j, i \in N$. And we build a center at location i if $1 \leq \hat{y}_i < 2$, i. e., set $\bar{y}_i := 1$ for each $i \in \hat{N}_1$. For $\hat{N}_2$, we will open at most $k - L_1$ locations as centers. If a center is not opened at location $j \in \hat{N}_2$, we assign the demand d'_j of j to another opened center i, i. e., set $\bar{x}_{ij} := 1$. Now we start to show the details of this step.

Initially, for each $i, j \in N$ set $\bar{x}_{ij} := 0$; and $\bar{y}_i := 0$. Then, we construct a collection of rooted trees spanning the locations in $\hat{N}_2$ as in [9]. Recall that $s(i)$ is the closest location to i in $(\hat{N}_1 \cup \hat{N}_2) - \{i\}$ $(N_1 \cup N_2 = \hat{N}_1 \cup \hat{N}_2)$ for each $i \in N_2$. We draw a directed edge from i to $s(i)$ if $i \in \hat{N}_2$. The cycles can be eliminated by the following way. For each cycle, we take any location in this cycle as a root and delete the edge from this root to other location. If there is a directed edge from i to $s(i)$ finally, we consider $s(i)$ as the parent of i. Then, we get a desired collection of rooted trees.

Next, we decompose each tree into a collection of rooted stars by Procedure 5.

▶ Remark. In each rooted star, all the children of the root have a fractional opening value. If the root of a star is a fractionally opened location, then the root has at least one child.

Procedure 5. Decompose a tree T to stars

while *there are at least two nodes in* T **do**
  choose a leaf node i with biggest number of edges on the path from i to the root;
  consider the subtree rooted at $s(i)$ as a rooted star, and remove this subtree;
end
if *only one node* i *is left and* $0 < \hat{y}_i < 1$ **then**
  add i to the star rooted at $s(i)$ as a child of $s(i)$;
end

▶ **Definition 11.** An even star is a star with even number of children. An odd star is a star with odd number of children.

Let Q_t denote the star rooted at location t. By abuse of notation, we also use Q_t to denote the collection of locations in the star rooted at t. Let $R_t = \sum_{i \in Q_t} \hat{y}_i$ be the total opening value in Q_t.

▶ **Lemma 12. (1)** *If a star* Q_t *has even positive number of fractionally opened locations, i. e.,* $|Q_t \cap \hat{N}_2| = 2q$ *is an even number and* $q \in \mathbb{Z}^+$*, then the total opening value of these fractionally opened locations is greater than* q*, i. e.,* $\sum_{i \in Q_t \cap \hat{N}_2} \hat{y}_i > q$.
(2) *If* $|Q_t \cap \hat{N}_2| = 2q + 1$ *is an odd number and* $q \in \mathbb{Z}^+$*, then* $\sum_{i \in Q_t \cap \hat{N}_2} \hat{y}_i > q + 1$.

Proof. (1) By property **3a**, $|\{i \in N \mid \frac{\alpha-2}{\alpha} < \hat{y}_i < \frac{\alpha-1}{\alpha}\}| \leq 1$. So in $\hat{N}_2$ at most one location has a fractional opening value in $(\frac{\alpha-2}{\alpha}, \frac{\alpha-1}{\alpha})$, and all other locations have fractional opening value exactly equal to $\frac{\alpha-1}{\alpha}$.
So,

$$\sum_{i \in Q_t \cap \hat{N}_2} \hat{y}_i > \frac{\alpha-2}{\alpha} + \frac{\alpha-1}{\alpha}(2q-1) = \frac{2q\alpha - 2q - 1}{\alpha} = q + \frac{q\alpha - 2q - 1}{\alpha}.$$

Moreover, since $\alpha \geq 4$ and $q \geq 1$, we have $\frac{q\alpha-2q-1}{\alpha} \geq \frac{2q-1}{\alpha} > 0$. Thus, $\sum_{i \in Q_t \cap \hat{N}_2} \hat{y}_i > q$.
(2) First, we have

$$\sum_{i \in Q_t \cap \hat{N}_2} \hat{y}_i > \frac{\alpha-2}{\alpha} + \frac{\alpha-1}{\alpha}2q = \frac{2q\alpha - 2q + \alpha - 2}{\alpha} = q + 1 + \frac{q\alpha - 2q - 2}{\alpha}.$$

Then, as $\alpha \geq 4$ and $q \geq 1$, we get $\frac{q\alpha-2q-2}{\alpha} \geq \frac{2q-2}{\alpha} \geq 0$. Thus, $\sum_{i \in Q_t \cap \hat{N}_2} \hat{y}_i > q + 1$. ◀

We build a center at each location $i \in \hat{N}_1 - \bigcup_t Q_t$ (locations are in $\hat{N}_1$, but not in any star), i. e., set $\bar{y}_i := 1$ and $\bar{x}_{ii} := 1$. For each kind of star Q_t, we define operations to make sure at most $\lfloor R_t \rfloor$ locations in Q_t are selected to be centers.

1. **An even star rooted at location** t **with** $1 \leq \hat{y}_t < 2$. Let $i_1, \cdots, i_{2q}$ be a sequence of all its children in nondecreasing order of distance from t. We build centers at location $t, i_1, i_3, \cdots, i_{2q-1}$, and serve the demand $d'_{i_{2r}}$ of i_{2r} by opened location i_{2r-1}, i. e.,

$$\text{set } \bar{y}_t := 1; \quad \bar{y}_{i_{2r-1}} := 1, \bar{y}_{i_{2r}} := 0, r = 1, \cdots, q;$$
$$\text{set } \bar{x}_{tt} := 1; \quad \bar{x}_{i_{2r-1}i_{2r-1}} := 1, \bar{x}_{i_{2r-1}i_{2r}} := 1, r = 1, \cdots, q.$$

2. **An even star rooted at location t with $\frac{\alpha-2}{\alpha} < \hat{y}_t \leq \frac{\alpha-1}{\alpha}$.** Let $i_1, \cdots, i_{2q}$ be a sequence of all its children in nondecreasing order of distance from t. (Note that $q \geq 1$ by the before Remark.) We build centers at location $t, i_2, i_4, \cdots, i_{2q}$, and serve the demand $d'_{i_{2r+1}}$ of i_{2r+1} by opened location i_{2r}, serve the demand d'_{i_1} of i_1 by t.
3. **An odd star rooted at location t with $1 + \frac{2}{\alpha} \leq \hat{y}_t < 2$.** Let $i_1, \cdots, i_{2q+1}$ be a sequence of all its children in nondecreasing order of distance from t. We open $t, i_1, i_3, \cdots, i_{2q+1}$ as centers, and serve the demand $d'_{i_{2r}}$ of i_{2r} by opened location i_{2r-1}.
4. **An odd star rooted at location t with $\frac{\alpha-2}{\alpha} < \hat{y}_t \leq \frac{\alpha-1}{\alpha}$ or $1 \leq \hat{y}_t < 1 + \frac{2}{\alpha}$.** Let $i_1, \cdots, i_{2q+1}$ be a sequence of all its children in nondecreasing order of distance from t. We build centers at location $t, i_2, i_4, \cdots, i_{2q}$, and serve the demand $d'_{i_{2r+1}}$ of i_{2r+1} by opened location i_{2r}, serve the demand d'_{i_1} of i_1 by t.

Note that $(\bar{x}, \bar{y})$ is an integral solution for new demands d'. To get an integral solution for our original demands d, we can redistribute the demands d' to their original locations according to Definition 9.

3 Analysis

By property **3a**, **3b** and **3c**, and Lemma 12, we can get the following lemma.

▶ **Lemma 13.** *For each kind of star Q_t, we build at most $\lfloor R_t \rfloor$ centers. And for each $i \in N$, we have $\sum_{j\in N} d'_j \bar{x}_{ij} \leq (2 + \frac{2}{\alpha}) M \bar{y}_i$.*

▶ **Lemma 14.** *We build at most k centers, and increase capacities by factor $2 + \frac{2}{\alpha}$.*

Proof. Suppose we get stars $Q_1, \cdots, Q_t$ by decomposing all the trees in Step 4. Then by property **3d**, we know $\sum_{r=1}^{t} R_r + \sum_{i\in \hat{N}_1 - \bigcup_{r=1}^{t} Q_r} \hat{y}_i \leq k$. Moreover, we build at most $\sum_{r=1}^{t} \lfloor R_r \rfloor + \sum_{i\in \hat{N}_1 - \bigcup_{r=1}^{t} Q_r} \lfloor \hat{y}_i \rfloor$ centers by Lemma 13 and the operation for locations that are in $\hat{N}_1$ but not in any star. Consequently, we build at most k centers. Again, by Lemma 13 we increase the capacity by at most a factor of $2+\frac{2}{\alpha}$ to satisfy all the demand constraints. ◀

For each location i in star Q_t, let $r(i) \in Q_t$ denote the location that the demand d'_i of i is reassigned to. Define the cost of star Q_t as $\sum_{i\in Q_t} d'_i c_{r(i)i}$.

▶ **Lemma 15.** *The cost of stars can be bounded by $\sum_{i\in N_2} \sum_{j\in N} \sum_{i'\in M_i} d_j (4c_{i'j} x_{i'j} + 8\alpha C_j x_{i'j})$.*

Proof. Note that in this proof we only consider location $i \in \hat{N}_2$, since we always build a center at each location in $\hat{N}_1$ and serve its demand by itself.

For each star Q_t, the reassignment is always to serve the demand d'_i of location i by an opened location i' that is closer to the root t, where $i, i' \in Q_t$ and $c_{ti'} \leq c_{ti}$. Recall that $s(i)$ is the closest location to i in $(N_1 \cup N_2) - \{i\}$. By Procedure 5, we know $s(i) = s(i') = t$. The cost for this reassignment is $d'_i c_{i'i}$, which can be bounded by $2d'_i c_{s(i)i}$ as $c_{i'i} \leq c_{s(i)i'} + c_{s(i)i} \leq 2c_{s(i)i}$.

Since $\frac{\alpha-2}{\alpha} < \hat{y}_i \leq \frac{\alpha-1}{\alpha}$ for each $i \in Q_t \cap \hat{N}_2$, we have $2d'_i c_{s(i)i} \leq 2\alpha(1 - \hat{y}_i) d'_i c_{s(i)i}$.

We sum $2\alpha(1 - \hat{y}_i) d'_i c_{s(i)i}$ over all $i \in \hat{N}_2$ to get an upper bound for the total cost of stars, i. e., $\sum_{i\in \hat{N}_2} 2\alpha(1 - \hat{y}_i) d'_i c_{s(i)i}$. Note that $\hat{N}_2 \subseteq N_2$. Then, by property **3e**,the definition of d'_i and Procedure 3 (Lemma 6), we know

$$\sum_{i\in \hat{N}_2} 2\alpha(1 - \hat{y}_i) d'_i c_{s(i)i} \leq \sum_{i\in N_2} \sum_{j\in N} \sum_{i'\in M_i} 2\alpha(1 - y'_i) d_j x_{i'j} c_{s(i)i}.$$

Therefore, it is sufficient to show that for each $j \in N, i' \in M_i, i \in N_2$

$$2\alpha(1-y_i')d_j x_{i'j} c_{s(i)i} \le d_j(4c_{i'j}x_{i'j} + 8\alpha C_j x_{i'j}).$$

We have two cases: (a) $N'(j) = i$ and (b) $N'(j) \neq i$. We show the above inequality holds for both cases.

(a) $N'(j) = i$.
Since $y_i' \in [\frac{\alpha-1}{\alpha}, 1), \forall i \in N_2$, we can find a location $i^* \notin M_i$ with $x_{i^*i} > 0$ and $c_{i^*i} \le \frac{C_i}{1-y_i'}$. Otherwise, $\sum_{r\in N} x_{ri}c_{ri} > C_i$, a contradiction.
Note that $c_{N'(i^*)i^*} \le c_{ii^*}$ since $N'(i^*) \neq i$, and $N'(i^*)$ is the closest location to i^* in N', and $i \in N'$. So, $c_{s(i)i} \le c_{iN'(i^*)} \le c_{N'(i^*)i^*} + c_{ii^*} \le 2c_{ii^*} \le 2\frac{C_i}{1-y_i'}$.
If $C_i \le C_j$, then we have

$$2\alpha(1-y_i')d_j x_{i'j} c_{s(i)i} \le 2\alpha d_j x_{i'j} 2C_i \le 4\alpha d_j x_{i'j} C_j. \tag{2}$$

Otherwise $C_i > C_j$. Then, we consider location j before i when we choose the cluster cores N', and j can not be a cluster core. This means there exists a location $r \in N'$ with $C_r \le C_j$ and $C_{rj} \le 2\alpha C_j$ before we check whether j should be chosen as a cluster core. So, $2\alpha C_i < c_{ri} \le c_{rj} + c_{ij} \le 2\alpha C_j + 2\alpha C_j = 4\alpha C_j$. That is, $C_i \le 2C_j$. Thus, for this case we have

$$2\alpha(1-y_i')d_j x_{i'j} c_{s(i)i} \le 2\alpha d_j x_{i'j} 2C_i \le 8\alpha d_j x_{i'j} C_j. \tag{3}$$

(b) $N'(j) \neq i$.
The proof for this case is similar as that in [7, 12]. First, we have

$$c_{s(i)i} \le c_{N'(j)i} \le c_{i'i} + c_{i'N'(j)} \le 2c_{i'N'(j)} \le 2(c_{i'j} + c_{N'(j)j}),$$

where $i' \in M_i$.
By property **1a**, $c_{N'(j)j} \le 2\alpha C_j$. So, $c_{s(i)i} \le 2c_{i'j} + 4\alpha C_j$.
Note that $0 < \alpha(1-y_i') \le 1$ as $1 > y_i' \ge \frac{\alpha-1}{\alpha}, i \in N_2$. Thus, we have

$$2\alpha(1-y_i')d_j x_{i'j} c_{s(i)i} \le 2d_j x_{i'j}(2c_{i'j} + 4\alpha C_j) = d_j(4c_{i'j}x_{i'j} + 8\alpha C_j x_{i'j}). \tag{4}$$

From inequalities (2), (3) and (4), we get

$$2\alpha(1-y_i')d_j x_{i'j} c_{s(i)i} \le d_j(4c_{i'j}x_{i'j} + 8\alpha C_j x_{i'j}).$$

◀

In our algorithm, we reassign the service twice: in Step 2 and Step 4. The cost of reassignment for Step 2 (Step 4) can be bounded by Lemma 6 (Lemma 15). Combining these two upper bounds, the total cost can be bounded by

$$\begin{aligned}
&\sum_{i\in N_2}\sum_{j\in N}\sum_{i'\in M_i} d_j(2c_{i'j} + 2\alpha C_j)x_{i'j} + \sum_{i\in N'-N_2}\sum_{j\in N}\sum_{i'\in M_i} d_j(3c_{i'j} + 4\alpha C_j)x_{i'j} \\
&+ \sum_{i\in N_2}\sum_{j\in N}\sum_{i'\in M_i} d_j(4c_{i'j}x_{i'j} + 8\alpha C_j x_{i'j}) \\
&\le \sum_{i\in N}\sum_{j\in N} d_j(6c_{ij} + 10\alpha C_j)x_{ij} = \sum_{j\in N} d_j(6C_j + 10\alpha C_j) = (6+10\alpha)C_{LP}.
\end{aligned}$$

Then combining with Lemma 7 and 14, we can prove the following theorem.

▶ **Theorem 16.** *For any $\alpha \ge 4$, there is a $(6+10\alpha)$-approximation algorithm for the hard uniform capacitated k-median problem with increasing the capacity by factor at most $2+\frac{2}{\alpha}$.*

4 Extent Our Algorithm to Solve Another Model

As mentioned in the introduction, the following model is also considered in some references for the capacitated k-median problem, where variable x_{ij} indicates the fraction of the demand of client j that is served by facility i, and y_i indicates if facility i is open. Let y_i take value one if facility i is open and value zero otherwise. We denote this model by CKL.

$$\begin{aligned} \min \quad & \sum_{i\in F}\sum_{j\in D} d_j c_{ij} x_{ij} \\ \text{subject to:} \quad & \sum_{i\in F} x_{ij} = 1, \quad \forall j \in D; \qquad \sum_{j\in D} d_j x_{ij} \le M y_i, \quad \forall i \in F; \\ & \sum_{i\in F} y_i \le k; \qquad\qquad 0 \le x_{ij} \le y_i, \quad \forall i \in F, j \in D, \\ & y_i \in \{0,1\}, \quad \forall i \in F. \end{aligned} \tag{5}$$

Replacing constraints (5) by $0 \le y_i \le 1, i \in F$, we get the LP-relaxation of CKL.

4.1 The Algorithm

Let (x^0, y^0) be an optimal solution to the LP-relaxation of CKL. For each facility $i \in F$, define a demand

$$d_i^1 = \sum_{j\in D} d_j x_{ij}^0.$$

To make use of the algorithm presented in Section 2, we set $N := F$. That is, each location $i \in N$ has a capacity M and demand d_i^1. Then, we get an instance of CKM considered in Section 2. Suppose we get an integral solution (x^1, y^1) for this constructed instance by the algorithm proposed in Section 2.

Then, we construct an integral solution (x^*, y^*) for the original instance of CKL by redistributing the demands $d_{i'}^1$ of location (facility) $i' \in N$ back to clients D. That is, set $y^* := y^1$; and set $x_{ij}^* := \sum_{i'\in N}(x_{ii'}^1 x_{i'j}^0)$, for each $i \in N = F, j \in D$.

4.2 Analysis

We only blow up the capacity once at the moment when we use the algorithm proposed in Section 2 to resolve the constructed instance. Theorem 16 states that this violation ratio is at most $2 + \frac{2}{\alpha}$. Thus, we have the following result.

▶ **Lemma 17.** *(x^*, y^*) is an integral solution for CKL with $\sum_{j\in D} d_j x_{ij}^* \le (2 + \frac{2}{\alpha}) M y_i^*$ for each $i \in F$, where $\alpha \ge 4$.*

▶ **Lemma 18.** *For any $\alpha \ge 4$, there is a $(13 + 20\alpha)$-approximation algorithm for CKL by increasing the capacity by factor $2 + \frac{2}{\alpha}$.*

Proof. Let $COST(\cdot,\cdot)$ be the total cost of solution $(\cdot,\cdot)$. Let OPT_{CKL} and OPT_{CKM} be the optimal objective value of our original instance and constructed instance respectively.

By the process to obtain the constructed instance, we have $OPT_{CKM} \le OPT_{CKL} + COST(x^0, y^0)$. Then,

$$\begin{aligned} & COST(x^*, y^*) \\ & \le COST(x^1, y^1) + COST(x^0, y^0) \le (6 + 10\alpha) OPT_{CKM} + COST(x^0, y^0) \\ & \le (6 + 10\alpha)(OPT_{CKL} + COST(x^0, y^0)) + COST(x^0, y^0) \le (13 + 20\alpha) OPT_{CKL}, \end{aligned}$$

where the first inequality holds according to the process to get the solution (x^*, y^*) and triangle inequalities; the second inequality follows by Theorem 16; the last inequality holds as $COST(x^0, y^0) \leq OPT_{CKL}$. ◀

Acknowledgements. We thank Dion Gijswijt for insightful discussions.

References

1. Aaron Archer, Ranjithkumar Rajagopalan, and David B. Shmoys. Lagrangian relaxation for the k-median problem: New insights and continuity properties. In Giuseppe Di Battista and Uri Zwick, editors, *ESA*, volume 2832 of *LNCS*, pages 31–42. Springer, 2003.
2. Vijay Arya, Naveen Garg, Rohit Khandekar, Adam Meyerson, Kamesh Munagala, and Vinayaka Pandit. Local search heuristic for k-median and facility location problems. In Jeffrey Scott Vitter, Paul G. Spirakis, and Mihalis Yannakakis, editors, *STOC*, pages 21–29. ACM, 2001.
3. Yair Bartal, Moses Charikar, and Danny Raz. Approximating min-sum k-clustering in metric spaces. In Jeffrey Scott Vitter, Paul G. Spirakis, and Mihalis Yannakakis, editors, *STOC*, pages 11–20. ACM, 2001.
4. Paul S. Bradley, Usama M. Fayyad, and Olvi L. Mangasarian. Mathematical programming for data mining: Formulations and challenges. *INFORMS Journal on Computing*, 11(3):217–238, 1999.
5. Jaroslaw Byrka, Krzysztof Fleszar, Bartosz Rybicki, and Joachim Spoerhase. A constant-factor approximation algorithm for uniform hard capacitated k-median. *CoRR*, abs/1312.6550, 2013.
6. Jaroslaw Byrka, Thomas Pensyl, Bartosz Rybicki, Aravind Srinivasan, and Khoa Trinh. An improved approximation for k-median, and positive correlation in budgeted optimization. *CoRR*, abs/1406.2951, 2014.
7. Moses Charikar. *Algorithms for clustering problems.* PhD thesis, Standford University, 2000.
8. Moses Charikar and Sudipto Guha. Improved combinatorial algorithms for the facility location and k-median problems. In *FOCS*, pages 378–388. IEEE Computer Society, 1999.
9. Moses Charikar, Sudipto Guha, Éva Tardos, and David B. Shmoys. A constant-factor approximation algorithm for the k-median problem (extended abstract). In Jeffrey Scott Vitter, Lawrence L. Larmore, and Frank Thomson Leighton, editors, *STOC*, pages 1–10. ACM, 1999.
10. Julia Chuzhoy and Yuval Rabani. Approximating k-median with non-uniform capacities. In *SODA*, pages 952–958. SIAM, 2005.
11. Dion Gijswijt and Shanfei Li. Approximation algorithms for the capacitated k-facility location problems. *CoRR*, abs/1311.4759, 2013.
12. Sudipto Guha. *Approximation algorithm for faciity location problems.* PhD thesis, Standford University, 2000.
13. Anil K. Jain and Richard C. Dubes. *Algorithms for clustering data.* Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1988.
14. Kamal Jain, Mohammad Mahdian, and Amin Saberi. A new greedy approach for facility location problems. In John H. Reif, editor, *STOC*, pages 731–740. ACM, 2002.
15. Kamal Jain and Vijay V. Vazirani. Approximation algorithms for metric facility location and k-median problems using the primal-dual schema and lagrangian relaxation. *J. ACM*, 48(2):274–296, 2001.
16. Madhukar R. Korupolu, C. Greg Plaxton, and Rajmohan Rajaraman. Analysis of a local search heuristic for facility location problems. *J. Algorithms*, 37(1):146–188, 2000.

17 Shi Li and Ola Svensson. Approximating k-median via pseudo-approximation. In Dan Boneh, Tim Roughgarden, and Joan Feigenbaum, editors, *STOC*, pages 901–910. ACM, 2013.

18 Jyh-Han Lin and Jeffrey Scott Vitter. epsilon-approximations with minimum packing constraint violation (extended abstract). In S. Rao Kosaraju, Mike Fellows, Avi Wigderson, and John A. Ellis, editors, *STOC*, pages 771–782. ACM, 1992.