Improved approximation algorithms for the facility location problems with linear/submodular penalty

Yu Li*[1,2], Donglei Du†[2], Naihua Xiu‡[1], and Dachuan Xu§[3]

[1]Department of Mathematics, School of Science, Beijing Jiaotong University, 3 Shangyuancun, Haidian District, Beijing 100044, P.R. China.

[2]Faculty of Business Administration, University of New Brunswick, Fredericton, NB Canada E3B 5A3.

[3]Department of Applied Mathematics, Beijing University of Technology, 100 Pingleyuan, Chaoyang District, Beijing 100124, P.R. China.

February 4, 2012

Abstract

We consider the *facility location problem with submodular penalty* (FLPSP) and the *facility location problem with linear penalty* (FLPLP), two extensions of the classical *facility location problem* (FLP). First, we introduce a general algorithmic framework for a class of covering problems with *submodular* penalty, extending the recent result of Geunes et al. [12] with *linear* penalty. This framework leverages existing approximation results for the original covering problems to obtain corresponding results for their submodular penalty counterparts. Specifically, any LP-based α-approximation for the original covering problem can be leveraged to obtain an $\left(1-e^{-1/\alpha}\right)^{-1}$-approximation algorithm for the counterpart with submodular penalty. Consequently, any LP-based approximation algorithm for the classical FLP (as a covering problem) can yield, via this framework, an approximation algorithm for the submodular penalty counterpart. Second, by exploiting some special properties of FLP, we present an LP rounding algorithm which has the currently best 2-approximation ratio over the previously best 2.50 by Hayrapetyan et al. [14] for the FLPSP and another LP rounding algorithm which has the currently best 1.5148-approximation ratio over the previously best 1.853 by Xu and Xu [29] for the FLPLP, respectively.

*Email: liyubjut@gmail.com.

†Email: ddu@unb.ca.

‡Email: nhxiu@center.njtu.edu.cn.

§Corresponding author. Email: xudc@bjut.edu.cn.

Keywords: Approximation algorithm · Facility location problem · LP rounding · Submodular function
Mathematics Subject Classification (2000) Primary 68W25; Secondary 90C27

1. Introduction

The classical *facility location problem* (FLP) is one of the most important models in combinatorial optimization with applications in operations research and computer science in general, and inventory management and supply chain management in particular. This problem is NP-hard and therefore much attention has been focused on designing approximation algorithms with good performance. Following the first constant-factor approximation algorithm by Shmoys et al. [24], there was a long list of work on designing improved approximation algorithms for this problem over the years. As a result, four basic schemes have been emerged in the design of these approximation algorithms, namely *LP-rounding* [3, 8, 19, 24, 26], *primal-dual* [17], *dual-fitting* [15, 16, 22], and *local search* [4, 13, 18]. These four competitive and complementary schemes possess different features. LP-rounding is *non-combinatorial* in nature mainly because of the resolving of the underlying LP-relaxation, while the other three are *combinatorial* in nature, hence offering faster approximation algorithms than LP rounding. However, the first scheme usually allows us to design algorithms with better approximation ratios compared to the other three—e.g., the currently best approximation ratio of 1.488 by Li [19] is obtained by combining LP-rounding and dual-fitting algorithm. Among the latter three, primal-dual can be adapted to solve variants of the classical FLP. Dual-fitting is essentially one special type of primal-dual methods, and usually offers better approximation ratio than a typical primal-dual, but is less robust. Local search is more powerful on the hard capacitated version of the FLP (cf. [32]).

On the impossibility of approximation, Guha and Khuller [13] show that the approximation ratio for FLP is at least 1.463, unless $NP \subseteq DTIME(n^{\log\log n})$, later strengthened to unless $P = NP$ by Sviridenko ([27]). For other variants of the FLP, we refer to [1, 2, 6, 10, 20, 21, 23, 25, 30, 31, 32] and the references therein.

In this work, we study two facility location problems with penalty, namely the *facility location problem with linear penalty* (FLPLP) and the more general *facility location problem with submodular penalty* (FLPSP), first introduced, as important extensions of the classical FLP, by Charikar et al. [5] and Hayrapetyan et al. [14], respectively. The FLPLP can be formally defined as follows. Consider a set $\mathcal{F}$ of facilities and a set $\mathcal{D}$ of clients. For every facility $i \in \mathcal{F}$ and every client $j \in \mathcal{D}$, there is a nonnegative opening cost f_i, a penalty cost p_j, and a connection cost c_{ij}. Unlike the regular FLP, the FLPLP does not require that every client must be served by some open facility. Instead, a client j can be either served by an open facility or rejected for service with penalty cost p_j. The problem is to open a subset of facilities such that each client $j \in \mathcal{D}$ is either assigned to an open facility or rejected with the objective to minimize the total cost, including the open, connection and linear penalty costs. We assume that the connection cost between clients and facilities is metric, i.e., $c_{ij} \leq c_{ij'} + c_{i'j'} + c_{i'j}$. The FLPSP is similar to the FLPLP except that the linear penalty cost is replaced by a general monotonically increasing submodular function. A set function $P : 2^{\mathcal{D}} \to \mathbb{R}_+$ is *submodular* if $P(X \cap Y) + P(X \cup Y) \leq P(X) + P(Y)$ for any subsets $X, Y \in 2^{\mathcal{D}}$. We assume that $P(\emptyset) = 0$.

For the FLPLP, four constant-factor approximation algorithms were known in the literature. Charikar et al. [5] gave a primal-dual 3-approximation algorithm. Xu and Xu [28, 29] presented an LP-rounding based $2 + e^{-1} \approx 2.736$-approximation algorithm, and a combinatorial 1.853-approximation algorithm by integrating primal-dual with local search technique. Recently, Geunes et al. [12] gave an improved LP-rounding based 2.056-approximation algorithm. In particular, Geunes et al. [12] presented an algorithmic framework which can convert any LP-based α-approximation for the classical FLP to an $\left(1 - e^{-1/\alpha}\right)^{-1}$-approximation algorithm for the counterpart with linear penalty.

For the FLPSP, three approximation algorithms were proposed in the literature. Hayrapetyan et al. [14] gave a simple LP-rounding based 2.50-approximation algorithm. Chudak and Nagano [7] gave a faster $(2.50+\epsilon)$-approximation algorithm by solving a convex relaxation rather than an LP relaxation of the FLPSP. Very recently, Du et al. [9] presented a primal-dual 3-approximation algorithm.

In summary, for the FLPLP, the best known approximation ratio is 1.853 [29], and the best known non-combinatorial ratio is 2.056 [12]. For the FLPSP, the best known combinatorial approximation ratio is 3 [9] and the best known non-combinatorial ratio is 2.50 [14].

The main contributions of this work are summarized as follows.

(i) We extend Geunes et al. [12]'s algorithmic framework for linear penalty to submodular penalty by showing that our framework can leverage any LP-based α-approximation to construct an $\left(1 - e^{-1/\alpha}\right)^{-1}$-approximation algorithm for the counterpart with submodular penalty.

(ii) Combining a novel LP-rounding technique with the JMS algorithm of [15, 16], along with the exploitation of the special properties of the submodular penalty function, we provide an improved 2-approximation algorithm for the FLPSP over the previously best approximation ratio 2.50 [14].

(iii) Note that for the FLPLP, the existing combinatorial ratio is better than the existing non-combinatorial ratio. This phenomena is generally "abnormal" and our third contribution corrects this "abnormality" by offering the currently best LP-rounding 1.5148-approximation algorithm which exploits the special properties of the linear penalty function.

There are intrinsic differences between the two LP-based algorithms for the FLPSP in (ii) and the FLPLP in (iii) due to the essential difference between linear and submodular functions. In general, linear penalty functions possess important properties not applicable to submodular penalty functions. Our algorithm and analysis indicate that the latter is substantially harder to approximate than the former because the techniques for the FLPLP, such as that by Byrka and Aardal [3], is not directly applicable to the FLPSP. To overcome this difficulty, our algorithm for the latter will exploit the special structure of the optimal LP relaxation solution of the FLPSP.

The rest of this paper is organized as follows. In Section 2, we first show that any LP-based α-approximation algorithm for the classical FLP can be leveraged to an $\left(1-e^{-1/\alpha}\right)^{-1}$-approximation algorithm for the submodular penalty counterpart. This result will serve as a concrete example in deriving a general framework for a class of covering problems with submodular penalty. And this framework also extends Geunes et al. [12]'s technique for the linear penalty model to the submodular case. In Sections 3 and 4, we present the 2-approximation algorithm for the FLPSP, and the 1.5148-approximation algorithm for the FLPLP, respectively.

We use the following notations throughout the paper: $n_f = |\mathcal{F}|$ and $n_c = |\mathcal{D}|$.

2 Algorithmic scheme for problems with submodular penalty

We will use the FLPSP as an example to show our algorithmic framework, which then will be extended to a more general class of covering problems.

2.1 An LP-rounding approximation algorithm for the FLPSP

The following LP relaxation for the FLPSP first appeared in Hayrapetyan et al. [14].

$$\begin{aligned} \min \quad & \sum_{i\in\mathcal{F}}\sum_{j\in\mathcal{D}} c_{ij}x_{ij} + \sum_{i\in\mathcal{F}} f_i y_i + \sum_{S\subseteq\mathcal{D}} P(S) z_S \\ \text{s.t.} \quad & \sum_{i\in\mathcal{F}} x_{ij} + \sum_{S\subseteq\mathcal{D}: j\in S} z_S \geq 1, \quad \forall j \in \mathcal{D}, \\ & x_{ij} \leq y_i, \quad \forall i\in\mathcal{F}, j\in\mathcal{D}, \\ & x_{ij}, y_i, z_S \geq 0, \qquad \forall i\in\mathcal{F}, j\in\mathcal{D}, S\subseteq\mathcal{D}, \end{aligned} \tag{2.1}$$

where $P(S)$ is a nondecreasing submodular function and $P(\emptyset) = 0$.

Now, we give an LP-rounding based approximation algorithm similar to that by Geunes et al. [12].

Algorithm 2.1

STEP 1. Solve the LP relaxation (2.1) to obtain an optimal fractional solution (x^*, y^*, z^*).

STEP 2. Construct a new variable z such that $z_j := 1 - \sum_{i\in\mathcal{F}} x^*_{ij}$.

STEP 3. Select parameter β uniformly at random from the interval $[0, \delta)$.

STEP 4. Reject the subset $S := \{j | z_j \geq \beta\}$, and pay the penalty cost $P(S)$.

STEP 5. Construct an instance of the classical FLP with the set of facilities $\mathcal{F}$, the set of clients $\mathcal{D}\setminus\mathcal{S}$ and the connection cost c_{ij} ($i \in \mathcal{F}$, and $j \in \mathcal{D}\setminus\mathcal{S}$). Then run the 1.488-approximation algorithm [19] for the instance and assign the clients in $\mathcal{D}\setminus S$ to the closest open facilities.

Lemma 2.2 *The expected penalty cost of the solution generated by Algorithm 2.1 is no more than* $\delta^{-1}\sum_{S\subseteq\mathcal{D}} P(S)z_S^*$.

Proof : Note that (x^*, y^*, z^*) is an optimal fractional solution, implying that $z_j = 1 - \sum_{i\in\mathcal{F}} x_{ij}^* \leq \sum_{S\subseteq\mathcal{D}:j\in S} z_S^*$. Note further $z \leq \sum_{S\subseteq\mathcal{D}} z_S^* I(S)$, where $I(S)$ is the incidence vector of subset S. We recall the definition of Lovász extension function $P'(\cdot)$ of any given submodular function $P(\cdot)$ along with its properties for later usage.

$$\begin{aligned} P'(z) = \max \quad & \sum_{j\in\mathcal{D}} \alpha_j z_j \\ \text{s.t.} \quad & \sum_{i\in S} \alpha_j \leq P(S), \quad \forall S\subseteq\mathcal{D}, \\ & \alpha_j \geq 0, \qquad \forall j\in\mathcal{D}. \end{aligned} \tag{2.2}$$

Property 1. $P'(I(S)) = P(S)$ and $P'(\mathbf{0}) = P(\emptyset)$.

Property 2. For any number $a \geq 0$ and any vector $w \geq \mathbf{0}$, $P'(aw) = aP'(w)$.

Property 3. $P'(\cdot)$ is an monotonically increasing convex function.

From these properties, it is easy to show that

$$P'(z) \leq P'\left(\sum_{S\subseteq\mathcal{D}} z_S^* I(S)\right) \leq \sum_{S\subseteq\mathcal{D}} z_S^* P'(I(S)) = \sum_{S\subseteq\mathcal{D}} P(S)z_S^*. \tag{2.3}$$

Without loss of generality, we assume that $z_1 \leq z_2 \cdots \leq z_{n_c}$, where $n_c = |\mathcal{D}|$, and moreover there are exactly ℓ nonzero and distinct numbers $z_{k_1} < z_{k_2} \cdots < z_{k_\ell}$ in $\{z_j\}$. Let $S_i = \{j\in\mathcal{D}|z_j \geq z_{k_i}\}$ $(i = 1\ldots,\ell)$. Then these sets satisfy that $S_\ell \subset \cdots \subset S_1$ and are all the ones that could be rejected by our algorithm. Let $z_{k_0} = 0$. We estimate the rejecting probability q_{S_i} for any subset S_i. When S_i is rejected, the random variable β must fall in the interval $(z_{k_{i-1}}, z_{k_i}]$, implying that $q_{S_i} \leq \delta^{-1}(z_{k_i} - z_{k_{i-1}})$ since β is from the interval $[0,\delta)$. From Lemma 3.1 in [11], we can construct an optimal solution $\widehat{\alpha} = (\widehat{\alpha}_1,\ldots,\widehat{\alpha}_{n_c})$, where

$$\widehat{\alpha}_j = P(\{z_j, z_{j+1}, ..., z_{n_c}\}) - P(\{z_{j+1}, z_{j+2}..., z_{n_c}\}),\ j = 1,\ldots,n_c.$$

The following claim is immediate.

Claim 2.3 *The optimal solution* $\widehat{\alpha}$ *satisfies* $\sum_{j\in S_i}\widehat{\alpha}_j = P(S_i)$ *for* $i = 1\cdots\ell$*;* $\sum_{j\in S}\widehat{\alpha}_j \leq P(S)$ *for any* $S\subseteq\mathcal{D}$*; and* $\sum_j^{n_c}\widehat{\alpha}_j = P(\{z_j, z_{j+1}..., z_{n_c}\})$*, for any* $j \leq n_c$.

Now we can bound the expected penalty cost as follows,

$$\begin{aligned} E[P] &= \sum_{i=1}^{l} P(S_i) q_{S_i} = \sum_{i=1}^{l} \left(\sum_{j \in S_i} \widehat{\alpha}_j \right) q_{S_i} \\ &= \sum_{j \in \mathcal{D}} \left(\sum_{i=1}^{k_j} q_{S_i} \right) \widehat{\alpha}_j = \frac{1}{\delta} \sum_{j \in \mathcal{D}} z_j \widehat{\alpha}_j \\ &\leq \frac{P'(z)}{\delta} \leq \frac{1}{\delta} \sum_{S \subseteq \mathcal{D}} P(S) z_S^*, \end{aligned}$$

where the last inequality is from (2.3). □

Theorem 2.4 *Setting $\delta = 1 - e^{-1/1.488}$, the approximation ratio of Algorithm 2.1 for the FLPSP is no more than $\left(1 - e^{-1/1.488}\right)^{-1} \leq 2.044$.*

Proof : Let F, C, and P be the open, connection and penalty costs of the solution obtained by Algorithm 2.1, respectively.

For the penalty cost, Lemma 2.2 implies that

$$E[P] \leq \frac{1}{\delta} \sum_{S \subseteq \mathcal{D}} P(S) z_S^*.$$

For the open and connection costs, given any β, the set of unrejected clients is

$$\mathcal{D} \setminus S = \left\{ j \in \mathcal{D} | 1 - \sum_{i \in \mathcal{F}} x_{ij}^* < \beta \right\},$$

implying that for any $j \in \mathcal{D} \setminus \mathcal{S}$,

$$\sum_{i \in \mathcal{F}} x_{ij}^* > 1 - \beta.$$

Therefore $(x^*/(1-\beta), y^*/(1-\beta))$ is a feasible solution to the LP relaxation of the classical FLP. Recall that the 1.488-approximation algorithm [19] called in Step 5 of Algorithm 2.1 is an LP-based algorithm. So

$$F + C \leq \frac{1.488}{1-\beta} \left(\sum_{i \in \mathcal{F}} \sum_{j \in \mathcal{D}} c_{ij} x_{ij}^* + \sum_{i \in \mathcal{F}} f_i y_i^* \right),$$

implying that

$$\begin{aligned} \mathbb{E}[F + C] &\leq \int_0^\delta \frac{1}{\delta} \frac{1.488}{1-\beta} d\beta \left(\sum_{i \in \mathcal{F}} \sum_{j \in \mathcal{D}} c_{ij} x_{ij}^* + \sum_{i \in \mathcal{F}} f_i y_i^* \right) \\ &= \frac{1.488}{\delta} \ln \left(\frac{1}{1-\delta} \right) \left(\sum_{i \in \mathcal{F}} \sum_{j \in \mathcal{D}} c_{ij} x_{ij}^* + \sum_{i \in \mathcal{F}} f_i y_i^* \right). \end{aligned}$$

Finally we have

$$\mathbb{E}[F+C+P] \leq \max\left\{\frac{1}{\delta}, \frac{1.488}{\delta}\ln\left(\frac{1}{1-\delta}\right)\right\}\left(\sum_{i\in\mathcal{F}}\sum_{j\in\mathcal{D}} c_{ij}x_{ij}^* + \sum_{i\in\mathcal{F}} f_i y_i^* + \sum_{S\subseteq\mathcal{D}} P(S)z_S^*\right).$$

Choosing $\delta = 1 - e^{-1/1.488}$, we get

$$\begin{aligned}\mathbb{E}[F+C+P] &\leq \frac{1}{1-e^{-1/1.488}}\left(\sum_{i\in\mathcal{F}}\sum_{j\in\mathcal{D}} c_{ij}x_{ij}^* + \sum_{i\in\mathcal{F}} f_i y_i^* + \sum_{S\subseteq\mathcal{D}} P(S)z_S^*\right)\\ &\leq 2.044\left(\sum_{i\in\mathcal{F}}\sum_{j\in\mathcal{D}} c_{ij}x_{ij}^* + \sum_{i\in\mathcal{F}} f_i y_i^* + \sum_{S\subseteq\mathcal{D}} P(S)z_S^*\right).\end{aligned}$$

□

We remark that there is a convex relaxation for the FLPSP presented by Chudak and Nagano [7],

$$\begin{aligned}\min \quad & \sum_{i\in\mathcal{F}}\sum_{j\subseteq\mathcal{D}} c_{ij}x_{ij} + \sum_{i\in\mathcal{F}} f_i y_i + P'(z) && (2.4)\\ \text{s.t.} \quad & \sum_{i\in\mathcal{F}} x_{ij} + z_j \geq 1, \quad \forall j\in\mathcal{D},\\ & x_{ij} \leq y_i, \quad \forall i\in\mathcal{F}, j\in\mathcal{D}, && (2.5)\\ & x_{ij}, y_i, z_j \geq 0, \qquad \forall i\in\mathcal{F}, j\in\mathcal{D},\end{aligned}$$

in which,

$$\begin{aligned}P'(z) = \max \quad & \sum_{j\in\mathcal{D}} \alpha_j z_j\\ \text{s.t.} \quad & \sum_{j\in S} \alpha_j \leq P(S), \quad \forall S\subseteq\mathcal{D},\\ & \alpha_j \geq 0, \qquad \forall j\in\mathcal{D}.\end{aligned}$$

It is not hard to see that in Algorithm 2.1 the solution (x^*, y^*, z) constructed in Step 2 is actually an optimal solution to (2.5). To avoid solving (2.1) with exponential number of variables, we can replace the first two steps in Algorithm 2.1 by just solving the convex relaxation (2.5) to obtain an new algorithm with the same approximation ratio except the extra term ε.

2.2 General rounding framework

We can extend our rounding technique in the previous section to the following more general model:

$$\begin{aligned}\min \quad & \varphi(w) + \sum_{S\subseteq\mathcal{D}} P(S)z_S\\ \text{s.t.} \quad & w_j + \sum_{S\subseteq\mathcal{D}: j\in S} z_S \geq 1, \quad \forall j\in\mathcal{D}, && (2.6)\\ & w_j, z_S \in \{0,1\}, \qquad \forall j\in\mathcal{D}, S\subseteq\mathcal{D}.\end{aligned}$$

This model captures a class of covering problems, in which the clients can receive service or be rejected, and the subset of clients S_{rej} rejected will incur a penalty cost $P(S_{rej})$. The binary vector w indicates whether the client j receives service or not. The term φ in the objective function satisfies Assumption 2.5 below. This new problem embeds a subproblem, denoted as $\phi(w)$.

Assumption 2.5 *There exists a function $\bar{\varphi}$: $[0,1]^{n_c} \longmapsto R_+$, such that*

1) $\bar{\varphi}$ is a lower bound on φ, i.e. $\bar{\varphi}(w) \le \varphi(w)$, for all $w \in [0,1]^{n_c}$;

2) for any fixed $w \in \{0,1\}^{n_c}$, we can efficiently find a solution to $\phi(w)$ of cost at most $\alpha\bar{\varphi}(w)$, where $\alpha \ge 1$;

3) the optimization problem

$$\begin{aligned} \min \quad & \bar{\varphi}(w) + \sum_{S \subseteq \mathcal{D}} P(S) z_S \\ s.t. \quad & w_j + \sum_{S \subseteq \mathcal{D}: j \in S} z_S = 1, \quad \forall j \in \mathcal{D}, \\ & w_j, z_S \in [0,1], \qquad \forall j \in \mathcal{D}, S \subseteq \mathcal{D}, \end{aligned} \tag{2.7}$$

can be solved efficiently.

In general the subproblem $\varphi(w)$ is NP-hard—e.g., for the FLPSP, $\phi(w)$ is the classical FLP. The above assumption will make it possible for us to design a constant-factor approximation algorithm for the general model. Note that (2.7) is a relaxation of (2.6), and becomes (2.1) in the case of the FLPSP, for example.

The following algorithm is an extension of Algorithm 2.1.

Algorithm 2.6

STEP 1. Solve the relaxation problem (2.7) to obtain an optimal fractional solution (w^*, z^*).

STEP 2. Select the parameter β uniformly at random from the interval $[0, \delta)$.

STEP 3. Reject the subset $S := \{j | 1 - w_j^* \ge \beta\}$, and pay the penalty cost $P(S)$. Construct variable $w \in \{0,1\}^{n_c}$ by setting $w := I(\mathcal{D} \setminus S)$.

STEP 4. Find a solution to the subproblem $\phi(w)$ and serve all the unrejected clients $\mathcal{D} \setminus S$.

We need one more assumption called *scaling* property [12].

Assumption 2.7 *The function $\bar{\varphi}$ satisfies the scaling property if*

$$\bar{\varphi}(w) \le \frac{1}{1-\beta}\bar{\varphi}(w^*), \quad \forall\ w^* \in [0,1]^{n_c}, \forall\ 0 \le \beta < 1.$$

Note that this assumption indeed holds in the FLPSP, because $(x^*/(1-\beta), y^*/(1-\beta))$ is a feasible solution to the relaxed problem.

With Assumptions 2.5 and 2.7, similar proofs to those in Lemma 2.2 and Theorem 2.4 lead to the following result:

Theorem 2.8 *Setting $\delta = 1 - e^{-1/\alpha}$, the approximation ratio of Algorithm 2.6 for (2.6) is no more than $\left(1 - e^{-1/\alpha}\right)^{-1}$.*

3 Improved 2-approximation algorithm for the FLPSP

In the previous section, we presented a general algorithmic frame for a class of covering problems with submodular penalty, but this frame can be too crude to be applied to certain specific problems and hence sometimes we need to refine the general frame to yield improved approximation ratios. For example, when we applied it to the FLPSP earlier in Section 2.1, the algorithm and its analysis ignored the possibility where unrejected clients may have paid fractional cost for penalty. In this section, we will incorporate this ignored cost into the design and analysis of our algorithms to obtain improved approximation ratio for the FLPSP.

Algorithm 3.1

STEP 1. Solve the convex relaxation (2.4) of the problem to obtain an optimal fractional solution (x^*, y^*, z^*). Or solve the LP relaxation (2.1) and convert its solution into (x^*, y^*, z^*) for (2.4)

STEP 2. Reject the subset $S_r := \{j | z_j \geq \frac{1}{2}\}$, and pay the penalty cost $P(S)$.

STEP 3. Construct an instance of classical FLP with the set of facilities $\mathcal{F}$, the set of clients $\mathcal{D} \backslash S_r$ and the connection cost c_{ij}, $i \in \mathcal{F}$, $j \in \mathcal{D} \backslash S_r$. Then run the JMS algorithm [15, 16] for the instance and assign the clients in $\mathcal{D} \setminus S_r$ to the closest open facilities. □

We will show that Algorithm 3.1 offers an improved approximation ratio. After the Step 1 of Algorithm 3.1, we know the total cost of the optimal fractional solution (x^*, y^*, z^*) for (2.1) is

$$\sum_{i \in \mathcal{F}} \sum_{j \subseteq \mathcal{D}} c_{ij} x_{ij}^* + \sum_{i \in \mathcal{F}} f_i y_i^* + P'(z^*).$$

Moreover, we have

Lemma 3.2

(1) $\sum_{j \in \mathcal{D}} \alpha_j^* z_j^* = P'(z^*)$;

(2) $P(S_r) \leq 2 \sum_{j \in S_r} \alpha_j^* z_j^*$;

(3) For any $j \notin S_r$, if $z_j^* > 0$ and $x_{ij}^* > 0$, then $\alpha_j^* \geq c_{ij}$.

Proof : (1) and (2) follow from Claim 2.3 directly. So we focus on (3). Suppose on the contrary, there exists a client $j \notin S_r$ satisfying $z_j^* > 0$ and $x_{ij}^* > 0$ for some facility i, but $\alpha_j^* < c_{ij}$. Then we construct a new feasible solution to convex relaxation (2.4): $(x, y, z) = (x^* - \epsilon, y^*, z^* + \epsilon)$, where $\epsilon > 0$ is sufficiently small to maintain the order of $\{z^*\}$, i.e. $z_1 \leq z_2 \cdots z_{j-1} \leq z_j + \epsilon \leq z_{j+1} \cdots \leq z_{n_c}$. Then $\{\alpha^*\}$ is still the optimal solution of (2.2) and we get a better solution to the convex relaxation (2.4), a contradiction to the optimality of (x^*, y^*, z^*). □

We now prove the approximation ratio of the above algorithm.

Theorem 3.3 *Algorithm 3.1 is a 2-approximation algorithm for the FLPSP.*

Proof : For all $j \notin S_r$, we can construct a feasible fractional solution from (x^*, y^*) as follows: scale all the y^*'s by 2, and reassign the clients $j \notin S_r$ to its closest fractional open facilities to get a new fractional solution $(\bar{x}, \bar{y})$. Lemma 3.2 implies that the open and connection costs of the new fractional solution are respectively given by

$$\sum_{i \in \mathcal{F}} f_i \bar{y}_i = 2 \sum_{i \in \mathcal{F}} f_i y_i^*$$

and

$$\sum_{i \in \mathcal{F}} \sum_{j \in \mathcal{D} \setminus S_r} c_{ij} \bar{x}_{ij} \leq \sum_{i \in \mathcal{F}} \sum_{j \in \mathcal{D} \setminus S_r} c_{ij} x_{ij}^* + \sum_{j \in \mathcal{D} \setminus S_r} \alpha_j^* z_j^*.$$

Since the JMS algorithm [15, 16] has a bi-factor of $(1, 2)$, the cost for the unrejected clients is no more than

$$2 \left(\sum_{i \in \mathcal{F}} f_i y_i^* + \sum_{i \in \mathcal{F}} \sum_{j \in \mathcal{D} \setminus S_r} c_{ij} x_{ij}^* + \sum_{j \in \mathcal{D} \setminus S_r} \alpha_j^* z_j^* \right).$$

Combining with Lemma 3.2, the total cost of the solution generated by Algorithm 3.1 is no more than

$$2 \left(\sum_{i \in \mathcal{F}} \sum_{j \in \mathcal{D}} c_{ij} x_{ij}^* + \sum_{i \in \mathcal{F}} f_i y_i^* + P'(z^*) \right).$$

□

4 Improved 1.5148-approximation algorithm for the FLPLP

Recently, Li [19] offered an improved analysis of the LP rounding algorithm by Byrka and Aardal [3] to obtain the currently best approximation ratio for the classical to FLP. In this section, we extend the LP rounding algorithm by Byrka and Aardal and the analysis by Li to obtain an improved algorithm for the FLPLP, which has an natural integer programming formulation:

$$\begin{aligned}\min \quad & \sum_{i\in\mathcal{F}}\sum_{j\in\mathcal{D}} c_{ij}x_{ij} + \sum_{i\in\mathcal{F}} f_i y_i + \sum_{j\in\mathcal{D}} p_j z_j \\ \text{s.t.} \quad & \sum_{i\in\mathcal{F}} x_{ij} + z_j \geq 1, \quad \forall j \in \mathcal{D}, \\ & x_{ij} \leq y_i, \quad \forall i \in \mathcal{F}, j \in \mathcal{D}, \\ & x_{ij}, y_i, z_j \in \{0,1\}, \qquad \forall i \in \mathcal{F}, j \in \mathcal{D},\end{aligned} \tag{4.8}$$

where the binary variable x_{ij} indicates whether client j is connected to facility i or not, binary variable y_i indicates whether facility i is open or not, and binary variable z_j indicates whether client j is penalized or not.

Denote the optimal solution of the LP relaxation as (x^*, y^*, z^*) and the corresponding optimal dual solution as (α^*, β^*). The total optimal fractional cost includes three parts: the opening cost $F^* = \sum_{i\in\mathcal{F}} f_i y_i^*$, the connection cost $C^* = \sum_{i\in\mathcal{F}}\sum_{j\in\mathcal{D}} c_{ij}x_{ij}^* = \sum_{j\in\mathcal{D}} C_j^*$, and the penalty cost $P^* = \sum_{j\in\mathcal{D}} p_j z_j^* = \sum_{j\in\mathcal{D}} P_j^*$.

We have the following lemma to capture the property of the optimal fractional solution (x^*, y^*, z^*).

Lemma 4.1 *For any $j \in \mathcal{D}$, if $z_j^* > 0$ and $x_{ij}^* > 0$, then $p_j \geq c_{ij}$.*

Proof : Suppose on the contrary, there exists a client j satisfying $z_j^* > 0$ and $x_{ij}^* > 0$ for some facility i, but $p_j < c_{ij}$. Then we can generate a better feasible solution to the LP relaxation: $(x, y, z) = (0, y^*, z_j^* + x_{ij}^*)$, a contradiction to the optimality of (x^*, y^*, z^*). □

We partition the set of clients $\mathcal{D}$ into two subsets.

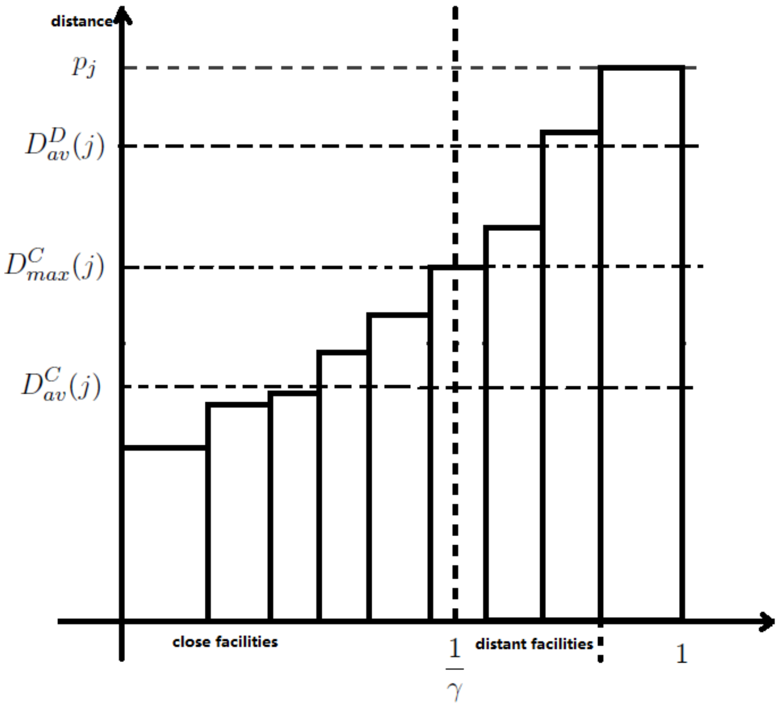
$$\begin{aligned}\mathcal{D}_\gamma &= \left\{ j \in \mathcal{D} | \sum_{i\in\mathcal{F}} x_{ij}^* \geq \frac{1}{\gamma} \right\}, \\ \bar{\mathcal{D}}_\gamma &= \mathcal{D} \setminus \mathcal{D}_\gamma.\end{aligned}$$

In our algorithm, we first randomly scale y^* by a number $\gamma > 1$ to obtain a suboptimal fractional solution $(x^*, \gamma y^*, z^*)$, making room for modification of x^* and z^* in order to reduce the total cost. Then, for each client j, we modify the corresponding x^* and z^* to reduce his fractional connection cost and his fractional penalty cost, resulting in a new feasible solution. However, Lemma 4.1 implies that for any client $j \in \mathcal{D}_\gamma$, the new z-variable must be zero. So for $j \in \mathcal{D}_\gamma$, we can omit z-variables and let $(\bar{x}, \bar{y})$ be the resultant *complete* solution (i.e. there exists no $i \in \mathcal{F}$ and $j \in \mathcal{D}$ such that $0 < \bar{x}_{ij} < \bar{y}_i$; otherwise we may split the facilities to obtain an equivalent instance with a complete solution—refer to Lemma 1 in [26] for a more detailed argument). For a client j, we say that a facility i is one of its *close* facilities if it fractionally serves client j in $(\bar{x}, \bar{y})$. If $\bar{x}_{ij} = 0$, but facility i was serving client j in solution (x^*, y^*, z^*), then we say that i is a *distant* facility of client j.

For any client $j \in \mathcal{D}$, let $i_1, i_2, \ldots, i_m$ be the facilities, arranged in the nondecreasing order of distances, to which client j is connected to fractionally in (x^*, y^*, z^*). Then we define $h_j(p) = c_{i_t,j}$, where t is minimum number such that $\sum_{s=1}^{t} y^*_{i_s} \geq p$, or $h_j(p) = p_j$ when $p > \sum_{s=1}^{m} y^*_{i_s}$.

Note that for every client $j \in \mathcal{D}$, the following facts hold:

- The sum of average connection and penalty costs equals $D_{av} = \int_0^1 h_j(p)\mathbf{d}p = F^*_j + P^*_j$.
- The average connection cost to a close facility equals $D^C_{av} = \gamma \int_0^{1/\gamma} h_j(p)\mathbf{d}p$.
- The average connection cost to a distant or rejected facility equals $D^D_{av} = \frac{\gamma}{\gamma-1} \int_{1/\gamma}^{1} h_j(p)\mathbf{d}p$.
- The maximal distance to a close facility is at most the average distance to a distant facility, $D^C_{\max} = h_j(1/\gamma)$.



We are now ready to present our algorithm. Consider the bipartite graph G obtained from the solution $(\bar{x}, \bar{y})$, where each client $j \in \mathcal{D}_\gamma$ is directly connected to his close facilities. Clients connected to the same facility in G are called *neighbors*. The main component in the algorithm is to cluster this graph recursively in a greedy fashion (Step 4), similarly to that used by Byrka and Aardal [3].

We let $\gamma = 1.3360$ with probability 0.45 and with the remaining 0.55 probability, we choose γ between (1.3360,1.9860] uniformly. We denote the distribution as $\mu(\gamma)$.

Algorithm 4.2

STEP 1. Solve the LP relaxation (4.8) to obtain an optimal fractional solution (x^*, y^*, z^*).

STEP 2. Scale up the value of the facility opening variables y^* by a random number γ obeying $\mu(\gamma)$. Then modify the value of the x^*, z^*-variables so that each client is connected to its closest fractionally open facilities.

STEP 3. If necessary, split facilities to obtain a complete solution $(\bar{x}, \bar{y})$.

STEP 4. Construct a greedy clustering from solution $(\bar{x}, \bar{y})$ by choosing recursively as cluster centers the unclustered clients in $\mathcal{D}_\gamma$ with minimal $D^C_{av}(j) + D^C_{\max}(j)$.

STEP 5. For every cluster center j, open one of its close facilities randomly with probabilities $\bar{x}_{ij}$.

STEP 6. For each facility i that is not a close facility of any cluster center, open it independently with probability $\bar{y}_i$.

STEP 7. For any client $j \in \mathcal{D}$, connect it to an open facility or pay penalty cost when the minimal connection cost to an open facility is larger than p_j.

We remark that only clients in $\mathcal{D}_\gamma$ can be chosen as cluster centers, which is different from the LP rounding algorithm in Byrka and Aardal [3] and Li [19].

During the analysis of the above algorithm, we will use the following lemma from [19] which holds also for our algorithm.

Lemma 4.3 *For any client j, we have*

$$E[C_j + P_j] \leq \int_0^1 h_j(p) e^{-\gamma p} \gamma \boldsymbol{d}p + e^{-\gamma} \left(\gamma \int_0^1 h_j(p) \boldsymbol{d}p + (3-\gamma) h_j \left(\frac{1}{\gamma} \right) \right).$$

Proof : Let us consider the following two possibilities.

Case 1. $j \in \mathcal{D}_\gamma$.

The lemma is concluded by Lemma 3 in [19].

Case 2. $j \in \bar{\mathcal{D}}_\gamma$.

We have

$$\begin{aligned} E[C_j + P_j] &\leq \int_0^1 h_j(p) e^{-\gamma p} \gamma \mathbf{d}p + e^{-\gamma} h_j \left(\frac{1}{\gamma} \right) \\ &\leq \int_0^1 h_j(p) e^{-\gamma p} \gamma \mathbf{d}p + e^{-\gamma} \left(\gamma \int_0^1 h_j(p) \mathbf{d}p + (3-\gamma) h_j \left(\frac{1}{\gamma} \right) \right), \end{aligned}$$

where the last inequality holds since $\gamma \leq 2$ in our distribution $\mu(\gamma)$. □

Now we present our main result for the FLPLP which follows the approach of Li [19]. We only give the sketch of the proof.

Theorem 4.4 *Algorithm 4.2 produces a solution with expected cost*

$$\mathbb{E}[F + C + P] \leq 1.5148(F^* + C^* + P^*),$$

implying that Algorithm 4.2 is an 1.5148*-approximation algorithm.*

Proof : The scaling factor for the facility cost is

$$\lambda_f = 0.45 \times 1.3360 + 0.55\frac{1.3360 + 1.9860}{2} = 1.5148.$$

To obtain the worst approximation factor λ_Q for the connection cost and penalty cost, we only need to consider step-functions as follows (since $h(p)$ is a nondecreasing function, which is a convex combination of step-functions, and the objective is a linear function):

$$h_q(p) = \begin{cases} 0, & \text{for } p < q, \\ \frac{1}{1-q}, & \text{for } p \geq q. \end{cases}$$

Now we can explicitly calculate $\lambda_Q(q)$ in three cases:

Case 1. $0 \leq q < 1/1.9860$.

$$\lambda_Q(q) = \frac{0.8462(e^{-1.3360q} - e^{-1.9860q})}{(1-q)q} + \frac{0.1138}{1-q} + 0.45\frac{e^{-1.3360q}}{1-q} + 0.3309.$$

Case 2. $1/1.9860 \leq q \leq 1/1.3360$.

$$\begin{aligned} \lambda_Q(q) &= \frac{0.8462(e^{-1.3360q} - e^{-1.9860q})}{(1-q)q} + \frac{0.1138}{1-q} + 0.45\frac{e^{-1.3360q}}{1-q} \\ &\quad + 0.3309 + \frac{0.8462}{1-q}(0.0019 - (2 - 1/q)e^{-1/q}). \end{aligned}$$

Case 3. $1/1.3360 < q < 1$.

$$\lambda_Q(q) = \frac{0.8462(e^{-1.3360q} - e^{-1.9860q})}{(1-q)q} + \frac{-0.2246}{1-q} + 0.45\frac{e^{-1.3360q}}{1-q} + 0.3309.$$

For all three cases, the maximum value of $\lambda_Q(q)$ is 1.5146, achieved at $q = 0.7400$ (See Fig. 2). Therefore we get a solution with value no more than than $1.5148F^* + 1.5146(C^* + P^*) \leq 1.5148(F^* + C^* + P^*)$. □

Acknowledgements This work was partially done while the first author was a visiting doctorate student at the Faculty of Business Administration, University of New Brunswick and supported in part by NSERC grants 283103. The research of the second author is supported by the Natural Sciences and Engineering

Research Council of Canada (NSERC) grant 283103. The third author's research is supported by the National Basic Research Program of China (No. 2010CB732501). The fourth author's research is supported by NSF of China (No. 11071268), Beijing Natural Science Foundation (No. 1102001), and Scientific Research Common Program of Beijing Municipal Commission of Education (No. KM201210005033).

References

[1] Aardal, K.I., Chudak, F.A., Shmoys, D.B.: A 3-approximation algorithm for the k-level uncapacitated facility location problem. Information Processing Letters **72**, 161-167 (1999)

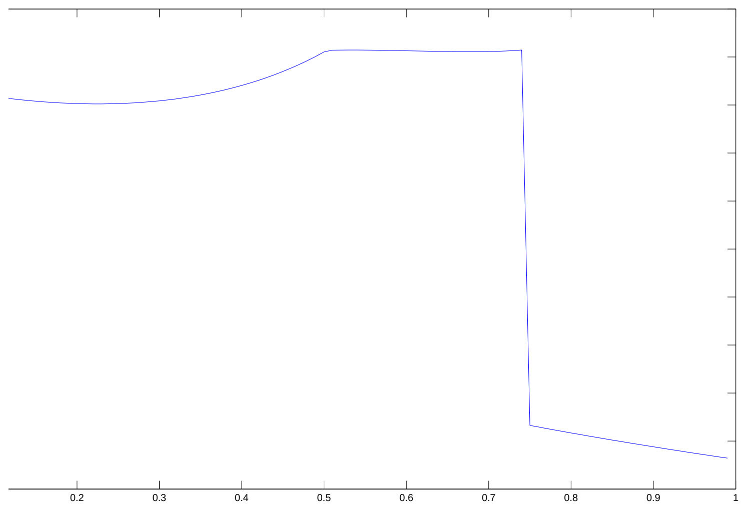
[2] Ageev, A., Ye, Y., Zhang, J.: Improved combinatorial approximation algorithms for the k-level facility location problem. SIAM Journal on Discrete Mathematics **18**, 207-217 (2003)

[3] Byrka, J., Aardal, K.I.: An optimal bi-factor approximation algorithm for the metric uncapacitated facility location problem. SIAM Journal on Computing **39**, 2212-2231 (2010)

[4] Charikar, M., Guha, S.: Improved combinatorial algorithms for facility location and k-median problems. In: Proceedings of FOCS, pp. 378-388 (1999)

[5] Charikar, M., Khuller, S., Mount, D.M., Naraasimban, G.: Algorithms for facility location problems with outliers. In: Proceedings of SODA, pp. 642-651 (2001)

[6] Chen, X., Chen, B.: Approximation algorithms for soft-capacitated facility location in capacitated network design. Algorithmica **53**, 263-297 (2007)



[7] Chudak, F.A., Nagano, K.: Efficient solutions to relaxations of combinatorial problems with submodular penalty via the Lovász extension and non-smooth convex optimization. In: Proceedings of SODA, pp. 79-88 (2007)

[8] Chudak, F.A., Shmoys, D.B.: Improved approximation algorithms for the uncapacitated facility location problem. SIAM Journal on Computing **33**, 1-25 (2003)

[9] Du, D., Lu, R., Xu, D.: A primal-dual approximation algorithm for the facility location problem with submodular penalty. Algorithmica **63**, 191-200 (2012)

[10] Du, D., Wang, X., Xu, D.: An approximation algorithm for the k-level capacitated facility location problem. Journal of Combinatorial Optimization **20**, 361-368 (2010)

[11] Fujishige, S.: Submodular Functions and Optimization (Second Edition), Annals of Discrete Mathematics, vol. 58, Elsevier (2005)

[12] Geunes, J., Levi, R., Romeijn, H.E., Shmoys, D.B.: Approximation algorithms for supply chain planning and logistics problems with market choice. Mathematical Programming **130**, 85-106 (2011)

[13] Guha, S., Khuller, S.: Greedy strike back: improved facility location algorithms. Journal of Algorithms **31**, 228-248 (1999)

[14] Hayrapetyan, A., Swamy, C., Tardös, E.: Network design for information networks. In: Proceedings of SODA, pp. 933-942 (2005)

[15] Jain, K., Mahdian, M., Markakis, E., Saberi, A., Vazirani, V.V.: Greedy facility location algorithms analyzed using dual fitting with factor-revealing LP. Journal of the ACM **50**, 795-824 (2003)

[16] Jain, K., Mahdian, M., Saberi, A., A new greedy approach for facility location problems. In: Proceedings of STOC, pp. 731-740 (2002)

[17] Jain, K., Vazirani, V.V.: Approximation algorithms for metric facility location and k-median problems using the primal-dual schema and Lagrangian relaxation. Journal of the ACM **48**, 274-296 (2001)

[18] Korupolu, M.R., Plaxton, C.G., Rajaraman, R.: Analysis of a local search heuristic for facility location problems. In: Proceedings of SODA, pp. 1-10 (1998)

[19] Li, S.: A 1.488 Approximation algorithm for the uncapacitated facility location problem. In: Proceedings of the ICALP, Part II, pp. 77-88 (2011)

[20] Li, Y., Shu, J., Wang, X., Xiu, N., Xu, D., Zhang, J.: Approximation algorithms for integrated distribution network design problems. Submitted (2010)

[21] Mahdian, M.: Facility location and the analysis of algorithms through factor-revealing programs. Ph. D. thesis, MIT, Cambridge, MA (2004)

[22] Mahdian, M., Ye, Y., Zhang, J.: Improved approximation algorithms for metric facility location problems. SIAM Journal on Computing **36**, 411-432 (2006)

[23] Shmoys, D.B., Swamy, C.: An approximation scheme for stochastic linear programming and its application to stochastic integer programs. Journal of the ACM **53**, 978-1012 (2006)

[24] Shmoys, D.B., Tardös, E., Aardal, K.I.: Approximation algorithms for facility location problems. In: Proceedings of STOC, pp. 265-274 (1997)

[25] Shu, J., Teo, C.P., Shen, Z.J. Max: Stochastic transportation-inventory network design problem. Operations Research **53**, 48-60 (2005)

[26] Sviridenko, M.: An improved approximation algorithm for the metric uncapacitated facility location problem. In: Proceedings of IPCO, pp. 240-257 (2002)

[27] Vygen, J.: Approximation algorithms for facility location problems (Lecture Notes). Report No. 05950-OR, Research Institute for Discrete Mathematics, University of Bonn, http://www.or.uni-bonn.de/vygen/fl.pdf (2005)

[28] Xu, G., Xu, J.: An LP rounding algorithm for approximating uncapacitated facility location problem with penalty. Information Processing Letters **94**, 119-123 (2005)

[29] Xu, G., Xu, J.: An improved approximation algorithm for uncapacitated facility location problem with penalty. Journal of Combinatorial Optimization **17**, 424-436 (2008)

[30] Ye, Y., Zhang, J.: An approximation algorithm for the dynamic facility location problem. Combinatorial Optimization in Communication Networks, Kluwer Academic Publishers, pp. 623-637 (2005)

[31] Zhang, J.: Approximating the two-level facility location problem via a quasi-greedy approach. Mathematical Programming **108**, 159-176 (2006)

[32] Zhang, J., Chen, B., Ye, Y.: A multiexchange local search algorithm for the capacitated facility location problem. Mathematics of Operations Research **30**, 389-403 (2005)