



# An $O(\log \log n)$ -Approximation for Submodular Facility Location

Fateme Abbasi  

University of Wrocław, Poland

Miguel Bosch-Calvo  

IDSIA, USI-SUPSI, Lugano, Switzerland

Fabrizio Grandoni  

IDSIA, USI-SUPSI, Lugano, Switzerland

Antoine Tinguely  

IDSIA, USI-SUPSI, Lugano, Switzerland

Marek Adamczyk  

University of Wrocław, Poland

Jarosław Byrka  

University of Wrocław, Poland

Krzysztof Sornat  

AGH University, Kraków, Poland

---

## Abstract

---

In the Submodular Facility Location problem (SFL) we are given a collection of  $n$  clients and  $m$  facilities in a metric space. A feasible solution consists of an assignment of each client to some facility. For each client, one has to pay the distance to the associated facility. Furthermore, for each facility  $f$  to which we assign the subset of clients  $S^f$ , one has to pay the opening cost  $g(S^f)$ , where  $g(\cdot)$  is a monotone submodular function with  $g(\emptyset) = 0$ .

SFL is APX-hard since it includes the classical (metric uncapacitated) Facility Location problem (with uniform facility costs) as a special case. Svitkina and Tardos [35, SODA'06] gave the current-best  $O(\log n)$  approximation algorithm for SFL. The same authors pose the open problem whether SFL admits a constant approximation and provide such an approximation for a very restricted special case of the problem.

We make some progress towards the solution of the above open problem by presenting an  $O(\log \log n)$  approximation. Our approach is rather flexible and can be easily extended to generalizations and variants of SFL. In more detail, we achieve the same approximation factor for the natural generalizations of SFL where the opening cost of each facility  $f$  is of the form  $p_f + g(S^f)$  or  $w_f \cdot g(S^f)$ , where  $p_f, w_f \geq 0$  are input values.

We also obtain an improved approximation algorithm for the related Universal Stochastic Facility Location problem. In this problem one is given a classical (metric) facility location instance and has to a priori assign each client to some facility. Then a subset of active clients is sampled from some given distribution, and one has to pay (a posteriori) only the connection and opening costs induced by the active clients. The expected opening cost of each facility  $f$  can be modelled with a submodular function of the set of clients assigned to  $f$ .

**2012 ACM Subject Classification** Theory of computation  $\rightarrow$  Approximation algorithms analysis; Theory of computation  $\rightarrow$  Facility location and clustering; Theory of computation  $\rightarrow$  Rounding techniques; Theory of computation  $\rightarrow$  Online algorithms

**Keywords and phrases** approximation algorithms, facility location, submodular facility location, universal stochastic facility location

**Digital Object Identifier** 10.4230/LIPIcs.ICALP.2024.5

**Category** Track A: Algorithms, Complexity and Games

**Related Version** *Full Version:* <https://arxiv.org/abs/2211.05474> [1]

**Funding** Fateme Abbasi and Jarosław Byrka were supported by Polish National Science Centre (NCN) Grant 2020/39/B/ST6/01641. Marek Adamczyk was supported by Polish National Science Centre (NCN) Grant 2019/35/D/ST6/03060. Miguel Bosch-Calvo, Fabrizio Grandoni, Krzysztof Sornat and Antoine Tinguely were supported by the SNSF Grant 200021\_200731/1. Krzysztof Sornat was partially supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 101002854).



© Fateme Abbasi, Marek Adamczyk, Miguel Bosch-Calvo, Jarosław Byrka, Fabrizio Grandoni, Krzysztof Sornat, and Antoine Tinguely; licensed under Creative Commons License CC-BY 4.0

51st International Colloquium on Automata, Languages, and Programming (ICALP 2024).

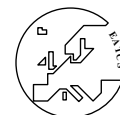
Editors: Karl Bringmann, Martin Grohe, Gabriele Puppis, and Ola Svensson;

Article No. 5; pp. 5:1–5:20



Leibniz International Proceedings in Informatics

LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany



**Acknowledgements** We would like to thank the anonymous reviewers for their helpful comments, in particular for pointing out to the simpler and stronger lower bound construction by Gupta [23]. We would also like to thank Neil Olver for inspiring discussions about applications of their technique in [5] to various covering problems over time.

## 1 Introduction

In the SUBMODULAR FACILITY LOCATION problem (SFL), we are given a set  $C$  of  $n$  clients and set  $F$  of  $m$  facilities, with metric distances  $d : (C \cup F) \times (C \cup F) \rightarrow \mathbb{R}_{\geq 0}$ . Furthermore, we are given a monotone submodular<sup>1</sup> (opening cost) function  $g : 2^C \rightarrow \mathbb{R}_{\geq 0}$  with  $g(\emptyset) = 0$ . Notice that  $g(\cdot)$  is non-negative. A feasible solution consists of an assignment  $\varphi : C \rightarrow F$  of each client to some facility (we also say that  $\varphi(c)$  serves  $c$ ). The opening cost of  $f \in F$  in this solution is  $g(\varphi^{-1}(f))$ . The cost of the solution, that we wish to minimize, is the sum of the distances from each client to the corresponding facility plus the total opening cost of the facilities, in other words

$$\text{cost}(\varphi) = \sum_{c \in C} d(c, \varphi(c)) + \sum_{f \in F} g(\varphi^{-1}(f)).$$

SFL captures practical scenarios where the cost of opening a facility is a (non-linear, still “tractable”) function of the set of served clients. For example, each client might have different types of needs, and satisfying such needs might have a submodular impact on the opening cost (regardless of the facility location). As we will discuss, SFL is also closely related to certain stochastic optimization problems which recently attracted a lot of attention (see, e.g. [2, 17, 19, 22, 24] and references therein). In particular, there are scenarios where one has to pay (a posteriori) the connection and opening costs related only to a random subset of *activated* clients, and this naturally induces objective functions with submodular opening costs.

SFL is APX-hard since it includes the classical FACILITY LOCATION problem (with uniform facility costs) as a special case [21]. Hence the best we can hope for, in terms of approximation algorithms, is a constant approximation. Finding such an approximation algorithm is explicitly posed as an open problem, e.g., by Svitkina and Tardos [35, 36]. The same authors present an  $O(\log n)$  approximation, based on a greedy approach, for a generalization of SFL where each facility  $f$  has a distinct monotone submodular function  $g_f(\cdot)$  (and this result is tight for this generalization due to a reduction from SET COVER by Shmoys, Swamy and Levi [33]). Chekuri and Ene [10] obtain an alternative  $O(\log n)$  approximation for the same generalization of SFL based on rounding a convex relaxation exploiting Lovász extensions (see also the related work on submodular partitioning problems [9, 13]). Svitkina and Tardos also present a constant approximation for a rather restrictive (still practically motivated) special case of SFL where  $g(\cdot)$  is induced by certain subtrees of a node-weighted tree over the clients.

### 1.1 Our Results and Techniques

We make some progress towards the resolution of the mentioned open problem by presenting an improved approximation algorithm for SFL.

<sup>1</sup> We recall that  $g(\cdot)$  is submodular iff, for every  $S, T \subseteq C$ ,  $g(S) + g(T) \geq g(S \cap T) + g(S \cup T)$ . The function is also monotone if  $g(T) \leq g(S)$  for every  $T \subseteq S \subseteq C$ . As usual in this framework, we assume to have an oracle access to  $g(\cdot)$ : given  $S \subseteq C$ , we can obtain the value of  $g(S)$  in polynomial time.

► **Theorem 1.1.** *There is a polynomial-time  $O(\log \log n)$ -approximation algorithm for SFL.*

Our approach is surprisingly simple (modulo exploiting some non-trivial results in the literature). By standard reductions (see Section 1.4) we can assume that  $N = n + m$  is polynomial in  $n$ , hence it is sufficient to provide an  $O(\log \log N)$  approximation. Our starting point is a natural (configuration) LP relaxation for the problem:

$$\begin{aligned}
\min \quad & \sum_{f \in F} \sum_{R \subseteq C} g(R) \cdot x_R^f + \sum_{c \in C} \sum_{f \in F} \sum_{R \ni c} d(c, f) \cdot x_R^f && \text{(Conf-LP)} \\
\text{s.t.} \quad & \sum_{f \in F} \sum_{R \ni c} x_R^f = 1 && \forall c \in C; \\
& \sum_{R \subseteq C} x_R^f = 1 && \forall f \in F; \\
& x_R^f \geq 0 && \forall R \subseteq C, \forall f \in F.
\end{aligned}$$

In an integral solution, we interpret  $x_R^f = 1$  as assigning exactly the set of clients  $R$  to the facility  $f$ . Notice that we impose  $\sum_{R \subseteq C} x_R^f = 1$ . This is w.l.o.g. since  $g(\emptyset) = 0$  (intuitively,  $x_\emptyset^f = 1$  means that no client is assigned to  $f$ ). We can solve the above LP in polynomial time (see Appendix A for a proof).

► **Lemma 1.2.** *In  $\text{poly}(N)$  time one can find an optimal solution to (Conf-LP) with  $\text{poly}(N)$  non-zero entries.*

Given an optimal solution  $\hat{x} = (\hat{x}_R^f)_{f \in F, R \subseteq C}$  to (Conf-LP) of cost  $\text{cost}(\hat{x})$  as in Lemma 1.2, we proceed with two main stages. In the first stage (discussed in Section 2) we simply sample partial assignments of clients to facilities with the distribution induced by  $\hat{x}$  for  $\ln \ln N$  many times. This cost at most  $\ln \ln N$  times the optimal LP cost in expectation, and leads to a partial solution that covers a random subset  $C_1 \subseteq C$  of clients.

In the second stage (discussed in Section 3) we take care of the remaining uncovered clients  $C_2 = C \setminus C_1$ . Let us consider the restriction  $\tilde{x}$  of  $\hat{x}$  to  $C_2$ . The opening cost of  $\tilde{x}$  might be as large as the opening cost of  $\hat{x}$ . However, in expectation, the connection cost of  $\tilde{x}$  is only a  $1/\ln N$  fraction of the connection cost of  $\hat{x}$  (as we will show).

At this point, using the probabilistic tree embedding algorithm in [14], we embed the original metric  $d$  into a (rooted) tree metric  $d^T$  over a hierarchically well-separated tree (HST)  $T$  (see Section 1.4 for the details). The opening cost of  $\tilde{x}$  w.r.t. to the new tree instance does not change, while its connection cost grows by a factor at most  $O(\log N)$  in expectation. Altogether we obtain a feasible fractional solution  $\tilde{x}$  over the tree instance whose expected cost is at most  $O(\text{cost}(\hat{x}))$ . Hence it is sufficient to develop an  $O(\log \log N)$ -approximate LP-rounding algorithm for the considered tree instance.

The next step is at the heart of our approach. Using the properties of HSTs and losing a constant factor in the approximation, we can further reduce our SFL tree instance to the following DESCENDANT-LEAF ASSIGNMENT problem (DLA): the facilities are leaves of  $T$  and the clients are arbitrary nodes of  $T$ . Each client  $c$  must be served by a facility contained in the subtree  $T_c$  rooted at  $c$ . The opening cost of each facility is given by  $g(\cdot)$ , and there are no connection costs at all. Bosman and Olver [5] describe a reduction of SUBMODULAR JOINT REPLENISHMENT and INVENTORY ROUTING problems to the NICE SUBADDITIVE COVER OVER TIME problem. We critically observed that DLA has some similarities with the latter problem (though this connection might not be obvious at first sight, see the discussion in Section 1.3). In particular, we were able to adapt their approach to achieve the desired  $O(\log \log N)$  approximation for our DLA problem.

## 5:4 An $O(\log \log n)$ -Approximation for Submodular Facility Location

We remark that we do not know how to get an  $O(1)$  approximation for SFL on trees (even on HSTs). Though such approximation would not imply an  $O(1)$  approximation for SFL with our approach (due to the first stage), finding it seems to be a natural intermediate problem to address.

The first stage of our construction might be helpful in other related problems, in particular to reduce the input problem to one on HSTs while introducing an additive  $O(\log \log n)$  term in the approximation ratio.

### 1.2 Generalizations and Variants

Our basic approach is rather flexible, and it can be applied to generalizations and variants of SFL. We next describe some other applications of our approach, and we expect to see a few more ones in the future. For example, we can handle the case where the opening cost of the facility  $f$  is  $g_f(S^f) = w_f \cdot g(S^f)$ , where  $w_f \geq 0$  is some input value: we call this the SFL WITH MULTIPLICATIVE OPENING COSTS problem (MULTSFL).

► **Theorem 1.3.** *There is a polynomial-time  $O(\log \log n)$ -approximation algorithm for MULTSFL.*

Similarly, we can address the SFL WITH ADDITIVE OPENING COSTS problem (ADDSFL), where  $g_f(S^f) = p_f + g(S^f)$  for  $S^f \neq \emptyset$ ,  $g_f(\emptyset) = 0$ , and  $p_f \geq 0$  is some input value.

► **Theorem 1.4.** *There is a polynomial-time  $O(\log \log n)$ -approximation algorithm for ADDSFL.*

The above generalizations are discussed in Appendix B. We remark that we do not know how to obtain an  $O(\log \log n)$ -approximation for the AFFINE SFL case, where the opening costs are submodular functions of the form  $g_f(S^f) = p_f + w_f \cdot g(S^f)$ . Notice that this generalizes both ADDSFL and MULTSFL. This is left as an interesting open problem.

As mentioned earlier, SFL is closely related to stochastic variants of FACILITY LOCATION. In particular, our approach also extends to the following UNIVERSAL STOCHASTIC FACILITY LOCATION problem (UNIVFL). Here we are given clients  $C$  and facilities  $F$  with metric distances  $d$  like in SFL, plus an opening cost  $w_f$  for each  $f \in F$ . Furthermore, we have an oracle access to a probability distribution  $\pi : 2^C \rightarrow \mathbb{R}_{\geq 0}$  specifying the probability  $\pi(A)$  that a given subset of clients  $A \subseteq C$  is activated. A feasible solution is an (universal) mapping  $\varphi : C \rightarrow F$ . The cost of  $\varphi$  w.r.t. clients  $A \subseteq C$  is  $\text{cost}_A(\varphi) = \sum_{c \in A} d(c, \varphi(c)) + \sum_{f \in F: \varphi^{-1}(f) \cap A \neq \emptyset} w_f$ . In words, this is the cost of connecting clients in  $A$  to the corresponding facilities, plus the cost of opening the facilities that serve at least one client in  $A$ . Our goal is to minimize  $\mathbb{E}_{A \sim \pi}[\text{cost}_A(\varphi)]$ . The main motivation for *universal* problems of this type is to allow a very quick (possibly distributed) reaction to requests that arrive over time. Let  $\text{opt} : C \rightarrow F$  minimize  $\mathbb{E}_{A \sim \pi}[\text{cost}_A(\text{opt})]$ , in other words  $\text{opt}$  is an optimal (universal) mapping. We say that an algorithm for UNIVFL is  $\alpha$ -approximate<sup>2</sup> if it returns a universal mapping  $\varphi$  satisfying  $\mathbb{E}_{A \sim \pi}[\text{cost}_A(\varphi)] \leq \alpha \cdot \mathbb{E}_{A \sim \pi}[\text{cost}_A(\text{opt})]$ .

Notice that the objective function of UNIVFL can be rewritten as

$$\sum_{c \in C} d(c, \varphi(c)) \cdot \mathbb{P}_{A \sim \pi}[\{c\} \cap A \neq \emptyset] + \sum_{f \in F} w_f \cdot \mathbb{P}_{A \sim \pi}[\varphi^{-1}(f) \cap A \neq \emptyset].$$

<sup>2</sup> In Section 1.3 we describe alternative ways to define the approximation ratio.

Hence UNIVFL is almost identical to SFL since  $g(R) = \mathbb{P}_{A \sim \pi}[R \cap A \neq \emptyset]$  is a monotone submodular function of  $R$  which is 0 for  $R = \emptyset$ . We can therefore adapt our techniques to achieve the following result (see Section 4). Let  $\pi_{\min} := \min_{c \in C} \{\mathbb{P}_{A \sim \pi}[c \in A]\}$  be the smallest probability of any client to be activated. W.l.o.g. we will assume  $\pi_{\min} > 0$ .

► **Theorem 1.5.** *There is a polynomial-time  $O(\log \log \frac{n}{\pi_{\min}})$ -approximation algorithm for the UNIVERSAL STOCHASTIC FACILITY LOCATION problem.*

For a comparison, Adamczyk, Grandoni, Leonardi and Włodarczyk [2] obtain an  $O(\log n)$  approximation which also holds for non-metric distances. In the case of metric distances, they obtain an  $O(1)$  approximation but only in the *independent activation case*, i.e., when the sampled set  $A$  of active clients is obtained by independently sampling each client  $c$  according to some input probability  $\pi'(c)$  for  $k$  times.

### 1.3 Related Work

As mentioned earlier, Bosman and Olver [5] consider the NICE SUBADDITIVE COVER OVER TIME problem: roughly speaking, here we are given a set  $V$  of items and a time interval  $\{1, \dots, L\}$ . Each item  $v \in V$  is associated with a time window  $F_v = \{s, \dots, t\}$ ,  $1 \leq s \leq t \leq L$ . The time windows altogether have a special *left-aligned* structure whose definition we skip here. A feasible solution consists of a subset  $S_t \subseteq V$  for each  $t \in \{1, \dots, L\}$ , such that, for each  $v \in V$ , one has  $v \in S_r$  for some  $r \in F_v$ . The goal is to minimize  $\sum_{t=1}^L g(S_t)$ , where  $g(\cdot)$  is a monotone submodular set function with  $g(\emptyset) = 0$ . For this problem they give a  $O(\log \log L)$  approximation, using a clever rounding algorithm for a convex optimization problem involving the Lovász extension of  $g(\cdot)$ . Intuitively, in our DLA problem (defined in Section 3.1) the time interval is replaced by the leaves (associated with some facility) of the tree  $\tilde{T}$ , and the time window of  $c \in \tilde{C}$  by the set  $\tilde{F}_c$ . Our time windows naturally induce a laminar family, which is a special case of the left-aligned structure mentioned before. The parameter  $\log L$  in their construction is replaced by the depth  $D$  of  $\tilde{T}$  in our case.

In the (METRIC UNCAPACITATED) FACILITY LOCATION problem (FL) we are given a set of clients and a set of facilities in a metric space  $d$ , where each facility has an opening cost  $o_f$ . One has to select a subset of facilities  $F' \subseteq F$  and assign each client  $c$  to the closest facility  $F'(c)$  in  $F'$  so as to minimize  $\sum_{c \in C} d(c, F'(c)) + \sum_{f \in F'} o_f$ . FL is a special case of both ADDSFL and MULTSFL (and of SFL in the case of uniform opening costs). FL is among the best-studied problems in the literature from the point of view of approximation algorithms (see, e.g., [8, 30, 34]). It is known to be APX-hard [21] and the current best-known 1.488-approximation algorithm [28] is a randomized combination of the greedy JMS algorithm [26] with an LP-rounding algorithm from [6]. Lagrangian-multiplier preserving algorithms for FL are at the heart of several approximation algorithms for fundamental clustering problems, including  $k$ -MEDIAN [3, 7, 11, 12, 18, 26, 27, 29] and  $k$ -MEANS [3, 11, 20].

Various variants of FL were studied in the literature and for most of them (at least with metric connection costs) a constant approximation was eventually discovered. A notable example is the CAPACITATED FACILITY LOCATION problem in which the number of clients that can be served from a facility is restricted by a location-specific bound. A local-search-based constant approximation for the latter problem is given in [37] (see also [4] for a more recent LP-based result). SFL is one of the most natural generalizations of (metric) FL where a constant approximation is still not known.

Grandoni, Gupta, Leonardi, Miettinen, Sankowski, and Singh [19], among other universal stochastic problems, studied UNIVFL in the independent activation case. However, they compare the cost of their solution with  $\mathbb{E}_{A \sim \pi}[\text{cost}_A(\text{opt}(A))]$ , where  $\text{opt}(A)$  is the optimal

facility location solution restricted to clients  $A$  (while we compare with  $\mathbb{E}_{A \sim \pi}[\text{cost}_A(\text{opt})]$ ). For this setting they obtain a  $O(\log n)$  approximation, which also holds for non-metric connection costs.

Gupta, Pál, Ravi, and Sinha [22] consider a 2-stage stochastic version of FL. Here in a first stage, one buys some facilities, then a subset of active clients is sampled from a given distribution. Finally, one can buy some more facilities, however at an opening cost which is increased by a multiplicative *inflation factor*  $\sigma$ . For this setting they present a constant approximation.

Universal stochastic problems have a natural online stochastic counterpart. For example, in the ONLINE STOCHASTIC FACILITY LOCATION problem clients are sampled one by one, and when client  $c$  is sampled one has to connect  $c$  to an already open facility or open a new facility  $f$  and connect  $c$  to  $f$ . Garg, Gupta, Leonardi and Sankowski [17] consider this problem in the independent activation case, i.e. when the next client to be served is sampled from a probability distribution  $\pi : C \rightarrow \mathbb{R}_{\geq 0}$ . For this setting, they present an  $O(1)$  approximation. Meyerson [31] studied a variant of the problem where an adversary chooses the set of input clients, and then a random permutation of them is presented in input (*random order model*).

We believe that it is plausible that SFL admits a constant approximation. In particular, one might consider greedy algorithms. Gupta [23] considered a natural set-cover type greedy algorithm for SFL. The same algorithm gives a 1.861-approximation when applied to the classical FACILITY LOCATION problem [26]. Gupta [23, Section 2.3] showed that this algorithm produces an  $\Omega(\log n)$  approximate solutions for SFL. Hence our algorithm is provably better than that one.

## 1.4 Preliminaries and Notation

We use  $\log$  for the logarithm with base 2 and  $\ln$  for the natural logarithm. Define  $X = C \cup F$ , and  $N = |X| = |C \cup F|$ . Given a metric  $d$  over  $X$ , we let  $d_{\min}$  be the smallest non-zero distance and  $d_{\max}$  be the largest distance (that we assume to be positive w.l.o.g). We use  $g(c)$  as a shortcut for  $g(\{c\})$ .

We sometimes express a feasible solution to SFL in the form  $S = (S^f)_{f \in F}$ , where  $S^f \subseteq C$  specifies the clients  $\varphi^{-1}(f)$  assigned to  $f$ . Notice that for each  $c \in C$  there is precisely one  $f \in F$  with  $c \in S^f$ . We define a *partial assignment* as  $S = (S^f)_{f \in F}$ , where  $S^f \subseteq C$ . We say that  $S$  covers the clients  $C' = \cup_{f \in F} S^f \subseteq C$ . Notice that, for technical reasons, in a partial assignment we allow  $S^f \cap S^{f'} \neq \emptyset$  for two distinct  $f, f' \in F$  (i.e. we allow to simultaneously assign a client to more than one facility). The cost of a (partial) assignment  $S$  of the above type is defined as  $\text{cost}(S) := \text{conn}(S) + \text{open}(S)$ , where  $\text{conn}(S) := \sum_{f \in F} \sum_{c \in S^f} d(c, f)$  is the connection cost of  $S$  and  $\text{open}(S) := \sum_{f \in F} g(S^f)$  is the opening cost of  $S$ . Given a (possibly infeasible) fractional solution  $x$  for (Conf-LP), we analogously define  $\text{cost}(x) = \text{conn}(x) + \text{open}(x)$ , where  $\text{conn}(x) = \sum_{c \in C} \sum_{f \in F} \sum_{R \ni c} d(c, f) \cdot x_R^f$ , and  $\text{open}(x) = \sum_{f \in F} \sum_{R \subseteq C} g(R) \cdot x_R^f$ .

It is convenient to define the merge  $S = S_1 + S_2$  of two partial assignments  $S_1$  and  $S_2$  naturally as follows: (1) for each facility  $f \in F$ , we initially set  $S^f := S_1^f \cup S_2^f$ ; (2) while there exist two distinct facilities  $f$  and  $f'$  with  $S^f \cap S^{f'} \neq \emptyset$ , replace  $S^{f'}$  with  $S^{f'} \setminus S^f$  (intuitively this second step guarantees that each client is assigned to no more than one facility). We observe that merging two partial assignments cannot increase the total cost.

► **Lemma 1.6.** *For any two partial assignments  $S_1$  and  $S_2$ ,  $\text{cost}(S_1 + S_2) \leq \text{cost}(S_1) + \text{cost}(S_2)$ .*

**Proof.** Let  $S = S_1 + S_2$ , and  $S'$  be the intermediate value of  $S$  obtained by executing only step (1) of the merge operation. One has  $\text{conn}(S') = \text{conn}(S_1) + \text{conn}(S_2)$ . Furthermore, by the submodularity (hence subadditivity) of  $g(\cdot)$ ,  $\text{open}(S') \leq \text{open}(S_1) + \text{open}(S_2)$ . Clearly  $\text{conn}(S) \leq \text{conn}(S')$ , and the monotonicity of  $g(\cdot)$  implies that  $\text{open}(S) \leq \text{open}(S')$ . The claim follows.  $\blacktriangleleft$

We will exploit the following fairly standard reductions (proofs in Appendix A), thanks to which in the following it will be sufficient to obtain an  $O(\log \log N)$  approximation for SFL. In order to distinguish between distinct instances  $J$  of the problem, we use  $\text{cost}_J(\varphi)$  to denote the cost of  $\varphi$  w.r.t.  $J$  and define similarly  $\text{open}_J(\varphi)$  etc.

► **Lemma 1.7.** *There is a 3-approximate reduction from SFL to the special case where  $m = n$ .*

► **Lemma 1.8.** *For any constant  $\varepsilon > 0$ , There is a  $(1 + 4\varepsilon)$ -approximate reduction from SFL to the special case where the metric  $d$  satisfies  $d_{\min} = 2$  and  $d_{\max} \leq \frac{2nN}{\varepsilon}$ .*

One of the key tools that we use is the notion of probabilistic tree embedding, which we use to map the input metric into a metric on a *hierarchically well-separated tree* (HST) while stretching the distances by a small enough factor. We recall that an HST is an edge weighted rooted tree where all the leaves are at the same distance from the root  $r$ . Furthermore, on every path from a leaf to  $r$  the edge weights are  $1, 2, 4, \dots$ . In particular, edges at the same level have the same weight. We will use the following construction<sup>3</sup> by Fakcharoenphol, Rao and Talwar [14].

► **Theorem 1.9** (FRT metric tree embedding [14]). *For any finite metric space  $(M, d)$  with  $d_{\min} > 1$ , there exists a randomized polynomial-time algorithm returning an HST  $T$  such that:*

1. *Every  $a \in M$  is mapped to some leaf  $\text{map}(a)$  of  $T$  (with elements at distance zero being mapped to the same leaf);*
2. *Let  $d^T(a, b) := d^T(\text{map}(a), \text{map}(b))$  be the length of the path between the leaves  $\text{map}(a)$  and  $\text{map}(b)$  of  $T$ . Then  $d^T(a, b) \geq d(a, b)$  and  $\mathbb{E}[d^T(a, b)] \leq 8 \log |M| \cdot d(a, b)$ ;*
3.  *$T$  has depth  $O(\log d_{\max})$ .*

For a given set  $C$ , let  $h: 2^C \rightarrow \mathbb{R}$  be a monotone submodular function with  $h(\emptyset) = 0$ . The *Lovász extension*  $\hat{h}: [0, 1]^C \rightarrow \mathbb{R}$  of  $h(\cdot)$  is defined as

$$\hat{h}(y) := \min \left\{ \sum_{R \subseteq C} h(R) \mu_R : \sum_{R \subseteq C} \sum_{R \ni c} \mu_R = y_c \ \forall c \in C, \sum_{R \subseteq C} \mu_R = 1, \mu \geq 0 \right\}. \quad (1)$$

The function  $\hat{h}(\cdot)$  is convex. We remark that  $\hat{h}(y)$  can be alternatively defined as

$$\hat{h}(y) := \sum_{k=1}^{n-1} h(\{c_1, \dots, c_k\}) (y_{c_k} - y_{c_{k+1}}) + h(C) y_{c_n}, \quad (2)$$

where the components of  $y$  are sorted in decreasing order, i.e.  $y_{c_1} \geq y_{c_2} \geq \dots \geq y_{c_n}$  [16, Section 6.3]. By the monotonicity of  $h(\cdot)$ ,  $\hat{h}(\cdot)$  is also non-decreasing in the sense that  $\hat{h}(y) \geq \hat{h}(y')$  if  $y \geq y'$ .

<sup>3</sup> We slightly and trivially extend their claim to consider nodes at distance 0.

## 2 Reducing the Connection Cost

In this section, we show how to compute a random partial assignment  $S_1 = (S_1^f)_{f \in F}$  covering a random subset of clients  $C_1 := \cup_{f \in F} S_1^f \subseteq C$  with the following high-level properties: the expected cost of  $S_1$  is “small enough” and (2) each client belongs to  $C_1$  with “large enough” probability. In the next section, we will describe a different partial assignment  $S_2 = (S_2^f)_{f \in F}$ , again of small enough cost, covering the remaining clients  $C_2 := C \setminus C_1$ . By merging these two partial assignments we obtain a feasible solution for the input problem of small enough total cost.

Let  $\hat{x}$  be an optimal solution to (Conf-LP) with at most  $\text{poly}(N)$  non-zero entries that can be computed via Lemma 1.2. The basic idea behind the next lemma is fairly standard: we sample partial assignments according to the distribution induced by  $\hat{x}$  for  $\ln \ln N$  times, and merge them together.

► **Lemma 2.1.** *In polynomial time one can compute a random partial assignment  $S_1$  covering a random subset of clients  $C_1$  such that: (1)  $\mathbb{E}[\text{cost}(S_1)] \leq \ln \ln(N) \cdot \text{cost}(\hat{x})$  and (2) For each  $c \in C$ ,  $\mathbb{P}[c \in C_1] \geq 1 - \frac{1}{\ln N}$ .*

**Proof.** For  $i \in \{1, 2, \dots, \ln \ln N\}$  and for every  $R \subseteq C$ , we define a partial assignment  $S(i, R)$  by setting  $S^f(i, R) = R$  independently with probability  $\hat{x}_R^f$  and  $S^f(i, R) = \emptyset$  otherwise. Let  $S_1 = \sum_{i=1}^{\ln \ln N} \sum_{R \subseteq C} S(i, R)$  be obtained by merging all these solutions, and let  $C_1 = \cup_{f \in F} S_1^f$ . Observe that

$$\mathbb{P}[c \notin C_1] = \prod_{f \in F} \prod_{R \ni c} (1 - \hat{x}_R^f)^{\ln \ln N} \leq e^{-\ln \ln N \sum_{f \in F} \sum_{R \ni c} \hat{x}_R^f} \leq e^{-\ln \ln N} = \frac{1}{\ln N}.$$

Furthermore, by Lemma 1.6,  $\mathbb{E}[\text{cost}(S_1)]$  is upper-bounded by

$$\sum_{i=1}^{\ln \ln N} \sum_{R \subseteq C} \mathbb{E}[\text{cost}(S(i, R))] = \ln \ln N \cdot \sum_{f \in F, R \subseteq C} \hat{x}_R^f \cdot \left( g(R) + \sum_{c \in R} d(c, f) \right) = \ln \ln N \cdot \text{cost}(\hat{x}). \quad \blacktriangleleft$$

Consider the partial assignment  $S_1$  covering the random subset of clients  $C_1$  as in the previous lemma. Let  $C_2 := C \setminus C_1$  be the remaining (uncovered) clients. Let also  $\tilde{x}$  be  $\hat{x}$  restricted to  $C_2$ , i.e.  $\tilde{x}_R^f = \sum_{R' \subseteq C_1} \hat{x}_{R \cup R'}^f$  for  $R \subseteq C_2$  and  $f \in F$ . The following lemma upper bounds the expected opening and connection cost of  $\tilde{x}$ .

► **Lemma 2.2.** *One has  $\text{open}(\tilde{x}) \leq \text{open}(\hat{x})$  and  $\mathbb{E}[\text{conn}(\tilde{x})] \leq \frac{1}{\ln N} \text{conn}(\hat{x})$ .*

**Proof.** We have  $\text{open}(\tilde{x}) \leq \text{open}(\hat{x})$  by the monotonicity of  $g(\cdot)$ . For the connection cost, notice that the probability of a client  $c$  being in  $C_2$  is at most  $1/\ln N$ , and only in that case one has to pay the associated connection cost. Thus by linearity of expectation, the expected connection cost of  $\tilde{x}$  is at most  $\text{conn}(\hat{x})/\ln N$ . The claim follows.  $\blacktriangleleft$

Notice that  $\tilde{x}$  is a feasible fractional solution for (Conf-LP) limited to  $C_2$ . In the following section, we show how to randomly round  $\tilde{x}$  to a partial assignment  $S_2$  which covers  $C_2$  at expected cost  $O(\log \log N) \cdot \text{cost}(\tilde{x})$ . It will then follow that  $S_1 + S_2$  is a feasible  $O(\log \log N)$ -approximate solution to the input SFL instance.

## 3 Approximating SFL on an HST

Given an SFL instance and a tree embedding of  $(C \cup F, d)$  into an HST  $T$  as in Theorem 1.9, we say that  $(C \cup F, d^T, g(\cdot), \text{map}(\cdot))$  is the corresponding HST-type instance. We remark that we allow multiple clients  $C(v)$  and facilities  $F(v)$  to be colocated at each leaf  $v$  of  $T$ . In this section we will describe an  $O(\log \log N)$ -approximate LP-rounding algorithm for the considered instances w.r.t. (Conf-LP).



► **Lemma 3.1.** *Given a feasible fractional solution  $x$  to (Conf-LP) for an HST-type SFL instance, in polynomial time one can compute a feasible (integral) solution for the same instance with cost at most  $O(\log \log N) \cdot \text{cost}(x)$ .*

Theorem 1.1 directly follows.

**Proof of Theorem 1.1.** By Lemma 1.7 it is sufficient to describe an  $O(\log \log N)$ -approximation. Furthermore by Lemma 1.8, we can assume that  $d_{\min} = 2$  and  $d_{\max} \leq \frac{2nN}{\epsilon}$ .

By applying the construction of Section 2 we compute a random partial assignment  $S_1 = (S_1^f)_{f \in F}$  covering the clients  $C_1 = \cup_{f \in F} S_1^f$  with expected cost at most  $O(\log \log N) \cdot \text{cost}(\hat{x})$ , where  $\hat{x}$  is an optimal solution to (Conf-LP). Furthermore, by Lemma 2.2, we obtain a feasible solution  $\tilde{x}$  to (Conf-LP) restricted to clients  $C_2 := C \setminus C_1$  which satisfies  $\text{open}(\tilde{x}) \leq \text{open}(\hat{x})$  and  $\mathbb{E}[\text{conn}(\tilde{x})] \leq \frac{1}{\ln N} \text{conn}(\hat{x})$ . By applying the probabilistic tree embedding from Theorem 1.9 to the metric  $(C_2 \cup F, d)$ , we obtain an HST-type SFL instance  $(C_2 \cup F, d^T, g(\cdot), \text{map}(\cdot))$  where the tree has depth  $D = O(\log d_{\max}) = O(\log N)$ . Observe that  $\tilde{x}$  is a feasible fractional solution for (Conf-LP) restricted to  $C_2$  on the HST-type instance. Furthermore, let  $\text{conn}_T(\tilde{x})$  denote the connection cost of  $\tilde{x}$  w.r.t. the HST-type instance, and define similarly  $\text{open}_T(\tilde{x})$  and  $\text{cost}_T(\tilde{x})$ . Then one has

$$\mathbb{E}[\text{cost}_T(\tilde{x})] = \text{open}(\tilde{x}) + \mathbb{E}[\text{conn}_T(\tilde{x})] \leq \text{open}(\hat{x}) + O(\log N) \cdot \mathbb{E}[\text{conn}(\tilde{x})] \leq O(\text{cost}(\hat{x})).$$

By applying the LP-rounding algorithm from Lemma 3.1 to  $\tilde{x}$  one obtains a partial assignment  $(S_2^f)_{f \in F}$  covering the clients  $C_2$  of cost at most  $O(\log \log N) \text{cost}(\hat{x})$ . The same solution has no larger cost in the original problem (on a non-tree metric). Altogether  $S_1 + S_2$  is a feasible solution to the input SFL problem of expected cost at most  $O(\log \log N) \cdot \text{cost}(\hat{x}) \leq O(\log \log N) \cdot \text{cost}(\text{opt})$ . ◀

In the rest of this section, we prove Lemma 3.1. To this aim, we will first present a reduction to a different problem that we call the DESCENDENT-LEAF ASSIGNMENT problem (DLA) (see Section 3.1). Then, we will present a good-enough approximation algorithm for DLA (see Section 3.2).

### 3.1 A Reduction to DLA

In the DESCENDENT-LEAF ASSIGNMENT problem (DLA) we are given a rooted tree  $\tilde{T}$  with depth  $D$ , a set of facilities  $\tilde{F}$  and a set of clients  $\tilde{C}$ . Each  $x \in \tilde{F} \cup \tilde{C}$  is mapped to some node  $v(x)$  of  $\tilde{T}$ , with the restriction that facilities are mapped to leaves of  $\tilde{T}$ . By  $\tilde{F}_c$  we denote the facilities which are mapped to nodes that are descendants of  $v(c)$  in  $\tilde{T}$  ( $v(c)$  included if it is a leaf). A feasible solution consists of an assignment  $\tilde{\varphi} : \tilde{C} \rightarrow \tilde{F}$  of each  $c \in \tilde{C}$  to some  $f \in \tilde{F}_c$ . The cost of this solution is  $\sum_{f \in \tilde{F}} h(\tilde{\varphi}^{-1}(f))$ , where  $h(\cdot)$  is a monotone submodular function over  $\tilde{C}$  with  $h(\emptyset) = 0$ . Similarly to SFL, we also express a feasible solution as  $S = (S^f)_{f \in \tilde{F}}$ , where  $S^f = \tilde{\varphi}^{-1}(f)$ , and let  $\text{cost}_{\text{DLA}}(S) = \sum_{f \in \tilde{F}} h(S^f)$  be the associated cost. We define a convex-programming (CP) relaxation for DLA as follows:

$$\begin{aligned} \min \quad & \sum_{f \in \tilde{F}} \hat{h}(z^f) && \text{(DLA-CP)} \\ \text{s.t.} \quad & \sum_{f \in \tilde{F}_c} z_c^f = 1 && \forall c \in \tilde{C}; \\ & z_c^f \geq 0 && \forall c \in \tilde{C}, \forall f \in \tilde{F}. \end{aligned}$$

## 5:10 An $O(\log \log n)$ -Approximation for Submodular Facility Location

In a 0-1 integral solution we interpret  $z_c^f = 1$  as  $c$  being assigned to  $f$ . Recall that  $\hat{h}(\cdot)$  is convex, which makes (DLA-CP) a convex program. We also notice that each feasible assignment  $S = (S^f)_{f \in \tilde{F}}$  corresponds to a feasible integral solution  $z = (z^f)_{f \in \tilde{F}}$  to (DLA-CP) with  $\text{cost}_{\text{DLA}}(S) = \text{cost}_{\text{DLA}}(z) := \sum_{f \in \tilde{F}} \hat{h}(z^f)$  and vice versa. Hence indeed (DLA-CP) is a CP-relaxation of DLA.

The next lemma provides the claimed reduction from SFL on HST-type instances to DLA.

► **Lemma 3.2.** *Given a polynomial-time  $O(\log D)$ -approximate CP-rounding algorithm for DLA w.r.t. (DLA-CP), where  $D$  is the depth of the tree, there is polynomial-time  $O(\log \log N)$ -approximate LP-rounding algorithm for SFL on HST-type instances with tree-depth  $O(\log N)$  w.r.t. (Conf-LP).*

**Proof.** Let  $(C \cup F, d^T, g(\cdot), \text{map}(\cdot))$  be the considered HST-type instance of SFL over an HST  $T$ , and  $x$  be an input feasible fractional solution to (Conf-LP) for this instance.

We build an instance  $(\tilde{C} \cup \tilde{F}, \tilde{T}, h(\cdot), v(\cdot))$  of DLA as follows. First, let  $y_c^f := \sum_{R \subseteq C: c \in R} x_R^f$ : intuitively this is the fractional amount by which  $c$  is assigned to  $f$  in  $x$ . We set  $h(\cdot) = g(\cdot)$  and  $\tilde{T} = T$ . Notice that  $D = O(\log N)$ . We set  $\tilde{F} = F$  and  $v(f) = \text{map}(f)$  for each  $f \in \tilde{F}$ . We associate to each  $c \in C$  a new client  $\tilde{c} \in \tilde{C}$ . Let  $T_v$  be the subtree rooted at  $v$  (containing  $v$  and all its descendants) and  $F_v$  be the facilities located in the leaves of  $T_v$  according to  $\text{map}(\cdot)$ . We map  $\tilde{c}$  into the lowest ancestor  $v(\tilde{c})$  of  $\text{map}(c)$  such that  $\sum_{f \in F_{v(\tilde{c})}} y_c^f \geq 1/2$ . Notice that  $v(\tilde{c}) = \text{map}(c)$  is possible (in which case there is at least one facility  $f$  colocated with  $c$  in  $T$ ).

We next define a feasible fractional solution  $z$  for (DLA-CP) w.r.t this DLA instance as follows. For each  $\tilde{c} \in \tilde{C}$  we set  $z_{\tilde{c}}^f = y_c^f / (\sum_{f' \in F_{v(\tilde{c})}} y_c^{f'})$  if  $f \in F_{v(\tilde{c})}$ , and otherwise  $z_{\tilde{c}}^f = 0$ . Let  $\tilde{\varphi}$  be a solution to the DLA instance obtained with the CP-rounding algorithm in the claim w.r.t.  $z$ . We obtain a feasible solution  $\varphi$  for the input instance by simply setting  $\varphi(c) = \tilde{\varphi}(\tilde{c})$ .

It remains to analyze the cost of  $\varphi$ . Define  $\bar{z}_{\tilde{c}}^f = y_c^f / (\sum_{f' \in F_{v(\tilde{c})}} y_c^{f'})$  for all  $f \in F$ . Notice that  $\bar{z} \geq z$ . By the definition of  $\hat{h}(\cdot)$  and its monotonicity,  $\hat{h}(z^f) \leq \hat{h}(\bar{z}^f) = \hat{h}(y_c^f / (\sum_{f' \in F_{v(\tilde{c})}} y_c^{f'})) \leq 2\hat{h}(y_c^f) = 2\hat{g}(y_c^f)$ . Notice that by plugging in  $x_R^f$  for  $\mu_R$  in the set in (1) and by how  $y$  is defined w.r.t.  $x$  above, we get  $\hat{g}(y_c^f) \leq \sum_{R \subseteq C} g(R) \cdot x_R^f$  and in particular  $\sum_{f \in F} \hat{g}(y_c^f) \leq \text{open}(x)$ . Thus, we have  $\text{cost}_{\text{DLA}}(z) \leq 2 \text{open}(x)$  and

$$\text{open}(\varphi) = \text{cost}_{\text{DLA}}(\tilde{\varphi}) = O(\log D) \cdot \text{cost}_{\text{DLA}}(z) \leq O(\log \log N) \cdot 2 \text{open}(x). \quad (3)$$

Consider next the connection cost of a given  $c \in C$ . If  $v(\tilde{c}) = \text{map}(c)$ , i.e  $v(\tilde{c})$  has no child, then  $d^T(c, \varphi(c)) = 0 \leq \sum_{f \in F} d^T(c, f) y_c^f$ . Otherwise, let  $w(c)$  be the child of  $v(\tilde{c})$  along the  $v(\tilde{c})$ - $\text{map}(c)$  path in  $T$ . Let  $\Delta$  be the weight of the edge between  $v(\tilde{c})$  and  $w(c)$ . Observe that the distance between  $v(\tilde{c})$  and the leaves in  $T_{v(\tilde{c})}$  is exactly  $2\Delta - 1$ . Furthermore, both  $c$  and  $\varphi(c)$  are located in the leaves of  $T_{v(\tilde{c})}$  in the HST mapping  $\text{map}(\cdot)$ . Hence  $d^T(c, \varphi(c)) \leq 2(2\Delta - 1)$ .

By the definition of  $v(\tilde{c})$ , it must be the case that  $\sum_{f \in F_{w(c)}} y_c^f < \frac{1}{2}$ , and consequently  $\sum_{f \in F \setminus F_{w(c)}} y_c^f \geq \frac{1}{2}$ . For each  $f \in F \setminus F_{w(c)}$ , the  $\text{map}(f)$ - $\text{map}(c)$  path in  $T$  has length at least  $2(2\Delta - 1)$ . Thus

$$\sum_{f \in F} d^T(c, f) y_c^f \geq \sum_{f \in F \setminus F_{w(c)}} d^T(c, f) y_c^f \geq \frac{1}{2} 2(2\Delta - 1).$$

Therefore, the connection cost of  $c$  in  $\varphi$  is at most 2 times its connection cost in  $x$ . We conclude that  $\text{conn}(\varphi) \leq 2 \text{conn}(x)$ . Altogether we have  $\text{cost}(\varphi) \leq 2 \text{conn}(x) + O(\log \log N) \cdot 2 \text{open}(x) \leq O(\log \log N) \cdot \text{cost}(x)$ .  $\blacktriangleleft$

### 3.2 An Approximation Algorithm for DLA

In this section, we present a CP-rounding algorithm for DLA. Lemma 3.1 follows by chaining Lemmas 3.2 and 3.3.

► **Lemma 3.3.** *Given a feasible fractional solution  $z$  to (DLA-CP) on an instance of DLA with tree-depth  $D$ , in polynomial time one can compute a feasible (integral) solution to the same instance of cost at most  $O(\log D) \cdot \text{cost}_{\text{DLA}}(z)$ .*

The CP-rounding algorithm from Lemma 3.3 is essentially the algorithm by Bosman and Olver [5] with minor modifications that we introduced to simplify our correctness analysis. Also, the analysis of its approximation ratio is essentially identical to [5], but we reproduce it for the sake of completeness. In particular, we will exploit the following definitions and lemma from [5]. Let  $h : 2^{\tilde{C}} \rightarrow \mathbb{R}_{\geq 0}$  be a monotone submodular function with  $h(\emptyset) = 0$ . For a given  $f \in \tilde{F}$  and a (possibly infeasible) solution  $z$  to (DLA-CP), let  $L_\theta(z^f) := \{c \in \tilde{C} : z_c^f \geq \theta\}$  be the set of clients that are served fractionally by at least some value  $\theta$  by  $f$ . Let also  $z^{f|\theta}$  be obtained from  $z^f$  by rounding down to  $\theta$  the values larger than  $\theta$ , i.e.  $z_c^{f|\theta} := \min\{z_c^f, \theta\}$  for each  $c \in \tilde{C}$ . Given  $\theta \in [0, 1]$  and  $z^f \in [0, 1]^{\tilde{C}}$ , we say that the set  $L_\theta(z^f)$  is  $\alpha$ -supported (w.r.t.  $h$ ) if  $\hat{h}(z^f) - \hat{h}(z^{f|\theta}) \geq \alpha h(L_\theta(z^f))$ .

► **Lemma 3.4** (Lemma 5.2 from [5]). *Given  $z^f \in [0, 1]^{\tilde{C}}$  and  $\alpha \in (0, 1]$ , at least one of the following holds: (1) there exists  $\theta \in [0, 1]$ , which can be computed in polynomial time, such that  $L_\theta(z^f)$  is  $\frac{\alpha}{32}$ -supported; (2)  $2^{1/\alpha} h(L_1(z^f)) \leq \hat{h}(z^f)$ .*

Our algorithm is Algorithm 1 in the figure. Recall that  $\tilde{T}_v$  is the subtree rooted at node  $v$ , where  $\tilde{T}_v$  includes  $v$  and all its descendants. Furthermore,  $\tilde{F}_v$  is the set of facilities mapped to the leaves of  $\tilde{T}_v$ . As usual the level of a node is its hop-distance from the root.

■ **Algorithm 1** An algorithm used to prove Lemma 3.3.

**Input:** Feasible solution  $z$  to (DLA-CP)

```

1:  $S^f \leftarrow \emptyset$  for all  $f \in F$ 
2: for  $i = 0, \dots, D$  do
3:   For every node  $v$  at level  $D - i$ , choose an arbitrary  $f_v \in \tilde{F}_v$  and set  $z^{f_v} \leftarrow \sum_{f' \in \tilde{F}_v} z^{f'}$  and
    $z^{f'} \leftarrow 0$  for all  $f' \in \tilde{F}_v \setminus \{f_v\}$ 
4:   if there exists  $\theta \in [0, 1]$  such that  $L_\theta(z^{f_v})$  is  $\frac{1}{32 \log(D+1)}$ -supported then
5:     For an arbitrary such  $\theta$ , set  $S^{f_v} \leftarrow S^{f_v} \cup L_\theta(z^{f_v})$  and  $z_c^{f_v} \leftarrow 0$  for all  $c \in L_\theta(z^{f_v})$ 
6:   else
7:     Set  $S^{f_v} \leftarrow S^{f_v} \cup L_1(z^{f_v})$  and  $z_c^{f_v} \leftarrow 0$  for all  $c \in L_1(z^{f_v})$ 
8:   For every  $c \in \tilde{C}$  choose  $f \in \tilde{F}_c$  such that  $c \in S^f$  and set  $S^{f'} \leftarrow S^{f'} \setminus \{c\}$  for all  $f' \in \tilde{F} \setminus \{f\}$ 
9: return  $(S^f)_{f \in \tilde{F}}$ 

```

Clearly Algorithm 1 runs in polynomial time. The next two lemmas analyze the correctness and the approximation ratio of Algorithm 1, hence proving Lemma 3.3.

► **Lemma 3.5.** *Algorithm 1 computes a feasible DLA solution.*

**Proof.** Consider a given client  $c \in \tilde{C}$  such that  $v(c)$  is at level  $D - i$  in  $\tilde{T}$ . Let us show that the following invariant holds at the beginning of each iteration  $j \leq i$ : either  $\sum_{f \in \tilde{F}_c} z_c^f = 1$  or  $c \in S^f$  for some  $f \in \tilde{F}_c$ . The invariant trivially holds for  $j = 0$ . Assume that it holds

## 5:12 An $O(\log \log n)$ -Approximation for Submodular Facility Location

up to the beginning of iteration  $j < i$ , and consider what happens during that iteration. Notice that for every node  $v$  at level  $D - j > D - i$ , we either have that every  $f \in \tilde{F}_v$  is a descendant of  $v(c)$  or every  $f \in \tilde{F}_v$  is not in  $\tilde{F}_c$ . Therefore, in Step (3) the value of  $\sum_{f \in \tilde{F}_c} z_c^f$  does not change. In more detail, it remains 1 by inductive hypothesis. The same value can decrease in Steps (5) or (7), however, this can only happen if  $c$  is added to  $S^{f_v}$  for some  $f_v \in \tilde{F}_c$ . Thus the invariant holds at the end of the  $j$ -th iteration, hence at the beginning of the next iteration  $j + 1$ .

Due to the invariant, during the iteration  $i$ , when one considers the node  $v = v(c)$ , one has that either  $c$  already belongs to some  $S^f$  with  $f \in \tilde{F}_c$ , or  $\sum_{f \in \tilde{F}_c} z_c^f = 1$ . In the latter case, after Step (3),  $z_c^{f_v} = 1$  where  $f_v \in \tilde{F}_c$ , so  $c$  belongs to every set  $L_\theta(z^{f_v})$  with  $\theta \in [0, 1]$ . As a consequence,  $c$  is added to  $S^{f_v}$  either in Step (5) or in Step (7).

It might happen that a client  $c$  is assigned *also* to a facility not in  $\tilde{F}_c$ . Step (8) guarantees that the final assignment of  $c$  is correct and unique.  $\blacktriangleleft$

► **Lemma 3.6.** *Algorithm 1 outputs a solution of cost at most  $O(\log D) \cdot \text{cost}_{\text{DLA}}(z)$ .*

**Proof.** Recall that  $\text{cost}_{\text{DLA}}(z) = \sum_{f \in \tilde{F}} \hat{h}(z^f)$ . We start by observing that the value of  $\text{cost}_{\text{DLA}}(z)$  can not increase over time when  $z$  changes during the execution of the algorithm. Indeed, Steps (5) and (7) can only decrease the entries of  $z$ , hence  $\text{cost}_{\text{DLA}}(z)$  by the monotonicity of  $\hat{h}(\cdot)$ . The only other changes of  $z$  happen in Step (3). Let us interpret this step as iteratively decreasing to zero  $z^{f'}$  for each  $f' \in \tilde{F}_v \setminus \{f_v\}$  and increasing  $z^{f_v}$  by the same amount. The decrease of the cost at each step is  $\hat{h}(z^{f_v}) + \hat{h}(z^{f'}) - \hat{h}(z^{f_v} + z^{f'})$ . By the alternative definition of  $\hat{h}(\cdot)$  as in (2) and its convexity, one has  $\hat{h}(z^{f_v} + z^{f'}) = 2\hat{h}\left(\frac{z^{f_v} + z^{f'}}{2}\right) \leq 2\left(\frac{1}{2}\hat{h}(z^{f_v}) + \frac{1}{2}\hat{h}(z^{f'})\right) = \hat{h}(z^{f_v}) + \hat{h}(z^{f'})$ . Hence the decrease of the cost is non-negative as required.

For each facility  $f$  and level  $i$ , let  $\Delta_i^\theta(f)$  be the clients added to  $S^f$  in Step (5) during iteration  $i$  (possibly  $\Delta_i^\theta(f) = \emptyset$ ). We define similarly  $\Delta_i^1(f)$  w.r.t. Step (7). Notice that, by the submodularity (hence subadditivity) of  $h(\cdot)$ , the increase of the cost of the solution due to adding  $\Delta$  to  $S^f$  is at most  $h(\Delta)$ . Therefore we can upper bound the cost of the final solution  $S = (S^f)_{f \in \tilde{F}}$  by

$$\text{cost}_{\text{DLA}}(S) := \sum_{f \in \tilde{F}} h(S^f) \leq \sum_{i=0}^D \sum_{f \in \tilde{F}} \left( h(\Delta_i^\theta(f)) + h(\Delta_i^1(f)) \right).$$

Let us upper bound the right-hand side of the above inequality. Let  $z(i)$  denote the value of  $z$  at the beginning of iteration  $i$ . From the previous observation, we have  $\hat{h}(z(i)) \leq \hat{h}(z)$  for every  $i$ . By Lemma 3.4 with  $\alpha = \frac{1}{\log(D+1)}$ , for any  $\Delta_i^1(f)$  one has  $h(\Delta_i^1(f)) \leq \frac{1}{D+1} \hat{h}(z^f(i))$ . Thus

$$\sum_{i=0}^D \sum_{f \in \tilde{F}} h(\Delta_i^1(f)) \leq \sum_{i=0}^D \sum_{f \in \tilde{F}} \frac{1}{D+1} \hat{h}(z^f(i)) \leq \sum_{i=0}^D \frac{1}{D+1} \text{cost}_{\text{DLA}}(z(i)) \leq \text{cost}_{\text{DLA}}(z). \quad (4)$$

Let  $z(D+1)$  be the value of  $z$  at the end of the  $D$ -th iteration, hence in particular  $\text{cost}_{\text{DLA}}(z(D+1)) \geq 0$ . Notice that  $z = z(0)$ . We can lower bound  $\text{cost}_{\text{DLA}}(z)$  by

$$\text{cost}_{\text{DLA}}(z) \geq \sum_{i=0}^D \left( \text{cost}_{\text{DLA}}(z(i)) - \text{cost}_{\text{DLA}}(z(i+1)) \right).$$

Let  $z_1(i)$  be the value of  $z$  obtained from  $z(i)$  after applying Step (3) for all nodes of level  $D - i$ . Let also  $z_2(i)$  be the value obtained from  $z_1(i)$  if, for all the facilities  $F'_i$  where Step (5) is applied during iteration  $i$ , instead of setting  $z_c^f = 0$  one sets  $z_c^f = \theta$  for the corresponding value of  $\theta$ . For the facilities not in  $F'_i$  we simply let  $z_2^f(i) = z_1^f(i)$ . Observe that  $z(i+1) \leq z_2(i) \leq z_1(i) \leq z(i)$ . One has

$$\begin{aligned} \text{cost}_{\text{DLA}}(z(i)) - \text{cost}_{\text{DLA}}(z(i+1)) &\geq \text{cost}_{\text{DLA}}(z_1(i)) - \text{cost}_{\text{DLA}}(z(i+1)) \\ &\geq \text{cost}_{\text{DLA}}(z_1(i)) - \text{cost}_{\text{DLA}}(z_2(i)) \\ &= \sum_{f \in \bar{F}} \hat{h}(z_1^f(i)) - \hat{h}(z_2^f(i)) = \sum_{f \in F'_i} \hat{h}(z_1^f(i)) - \hat{h}(z_2^f(i)) \\ &\geq \frac{\sum_{f \in F'_i} h(\Delta_i^\theta(f))}{32 \log(D+1)} = \frac{\sum_{f \in \bar{F}} h(\Delta_i^\theta(f))}{32 \log(D+1)}. \end{aligned}$$

In the first two inequalities above we used the monotonicity of  $\hat{h}(\cdot)$ , while in the last inequality the definition of  $\alpha$ -supported. Altogether

$$\begin{aligned} \sum_{i=0}^D \sum_{f \in \bar{F}} h(\Delta_i^\theta(f)) &\leq 32 \log(D+1) \cdot \sum_{i=0}^D (\text{cost}_{\text{DLA}}(z(i)) - \text{cost}_{\text{DLA}}(z(i+1))) \\ &\leq O(\log D) \cdot \text{cost}_{\text{DLA}}(z). \end{aligned} \quad (5)$$

By the monotonicity of  $h(\cdot)$ , Step (8) cannot increase the cost of the solution, hence the claim.  $\blacktriangleleft$

## 4 Universal Stochastic Facility Location

In this section we sketch our approximation algorithm for UNIVFL. We first present a weaker approximation factor  $O(\log \log N + \log \log \frac{d_{\max}}{d_{\min}})$ . Later we will show how to refine it.

Define  $g(R) := \mathbb{P}_{A \sim \pi}[R \cap A \neq \emptyset]$ . We observe that this function is monotone submodular and  $g(\emptyset) = 0$ . Recall that  $g(c) = g(\{c\})$  for every  $c \in C$ . W.l.o.g. we can assume  $g(c) > 0$  since otherwise we can discard  $c$ . We can define the objective function of UNIVFL for a given assignment  $\varphi : C \rightarrow F$  as

$$\text{cost}(\varphi) = \text{conn}(\varphi) + \text{open}(\varphi) = \sum_{c \in C} d(c, \varphi(c)) \cdot g(c) + \sum_{f \in F} w_f \cdot g(\varphi^{-1}(f)).$$

Notice that only the connection cost changes w.r.t. MULTSFL. In more detail, the connection cost of each client  $c$  is scaled by the factor  $g(c)$ .

We can similarly define a configuration LP for UNIVFL, and solve it by the same arguments as in Lemma 1.2. We next use an analogous notation as for SFL. Let  $\dot{x}$  be an optimal solution to this LP with  $\text{poly}(N)$  many non-zero variables. We can apply the first stage of our algorithm for SFL (described in Section 2) with essentially no changes. This will lead to a partial assignment  $S_1$  of expected cost  $\mathbb{E}[\text{cost}(S_1)] \leq \ln \ln N \cdot \text{cost}(\dot{x})$  and serving the clients  $C_1$ , where  $\mathbb{P}[c \notin C_1] \leq \frac{1}{\ln N}$ . Mapping the metric over an HST  $T$  and considering the restriction  $\ddot{x}$  of  $\dot{x}$  to  $C_2 := C \setminus C_1$ , we obtain that  $\mathbb{E}[\text{cost}_{\text{HST}}(\ddot{x})] = O(\text{cost}(\dot{x}))$ . A reduction similar to the one in Lemma 3.2 works also in this case (since the scaling of the fractional solution is done on a per-client base). However in this case  $D = O(\log \frac{d_{\max}}{d_{\min}})$  (since we did not reduce the ratio  $\frac{d_{\max}}{d_{\min}}$  in a preprocessing step). Hence we can apply the result from Lemma 3.3 to obtain an assignment covering  $C_2$  of expected cost  $O(\log \log \frac{d_{\max}}{d_{\min}}) \cdot \text{cost}(\dot{x})$ . This concludes the sketch of the  $O(\log \log N + \log \log \frac{d_{\max}}{d_{\min}})$  approximation.

We next improve this bound via a preprocessing step. Recall that  $0 < \pi_{\min} := \min_{c \in C} \{g(c)\}$ . We first scale the ratio  $d_{\max}/d_{\min}$ . Let us guess<sup>4</sup> the largest distance  $L = \max_{c \in C} \{d(c, \text{opt}(c))\}$  in some optimal (universal) solution  $\text{opt}$ . Notice that  $\text{cost}(\text{opt}) \geq \pi_{\min} L$ . We use essentially the same arguments as in Lemma 1.8, we can enforce that  $d_{\max} \leq NL$  and  $d_{\min} \geq \frac{\varepsilon}{n} \pi_{\min} L$ . Hence we obtain  $\frac{d_{\max}}{d_{\min}} \leq \frac{nN}{\varepsilon \pi_{\min}}$ .

Now let us reduce the number of facilities  $m$  to  $O(n + \log \frac{1}{\pi_{\min}})$  (hence  $N$  as well). Here we use essentially the same argument as in the proof of Lemma B.1 (with  $p_f = 0$ ). In more detail, we can assume that  $m \leq 2^n$ . Indeed, otherwise we can reduce the input instance to a WEIGHTED SET COVER instance (that we can solve exactly in polynomial time) in the same way as in the mentioned lemma, with the difference that now, for  $R \neq \emptyset$ , we set  $\kappa_R = \min_{f \in F} \{w_f \cdot g(R) + \sum_{c \in R} d(c, f) \cdot g(c)\}$ . By the rest of the construction in the same lemma, we can reduce (with a constant loss in the approximation factor) our instance to one where there are  $O(\log \frac{d_{\max}}{d_{\min}}) = O(\log \frac{n2^n}{\varepsilon \pi_{\min}}) = O(n + \log \frac{1}{\pi_{\min}})$  facilities per client. Altogether we reduce  $N$  to  $N' = O(n(n + \log \frac{1}{\pi_{\min}}))$ . Now we can apply again the above scaling trick over the distances (with  $N$  replaced by  $N'$ ) to obtain distances  $d'$  which satisfy:

$$\frac{d'_{\max}}{d'_{\min}} \leq \frac{nN'}{\varepsilon \pi_{\min}} = O\left(\frac{n^3 + n^2 \log \frac{1}{\pi_{\min}}}{\pi_{\min}}\right).$$

This leads to the approximation factor

$$O\left(\log \log \frac{d'_{\max}}{d'_{\min}} + \log \log N'\right) = O\left(\log \log \frac{n}{\pi_{\min}}\right).$$

---

## References

- 1 Fateme Abbasi, Marek Adamczyk, Miguel Bosch-Calvo, Jarosław Byrka, Fabrizio Grandoni, Krzysztof Sornat, and Antoine Tinguely. An  $O(\log \log n)$ -approximation for submodular facility location. *CoRR*, abs/2211.05474, 2022. doi:10.48550/arXiv.2211.05474.
- 2 Marek Adamczyk, Fabrizio Grandoni, Stefano Leonardi, and Michał Włodarczyk. When the optimum is also blind: A new perspective on universal optimization. In *44th International Colloquium on Automata, Languages, and Programming, ICALP 2017*, pages 35:1–35:15, 2017. doi:10.4230/LIPIcs.ICALP.2017.35.
- 3 Sara Ahmadian, Ashkan Norouzi-Fard, Ola Svensson, and Justin Ward. Better guarantees for  $k$ -means and Euclidean  $k$ -median by primal-dual algorithms. *SIAM J. Comput.*, 49(4), 2020. doi:10.1137/18M1171321.
- 4 Hyung-Chan An, Mohit Singh, and Ola Svensson. LP-based algorithms for capacitated facility location. *SIAM J. Comput.*, 46(1):272–306, 2017. doi:10.1137/151002320.
- 5 Thomas Bosman and Neil Olver. Improved approximation algorithms for inventory problems. In *Integer Programming and Combinatorial Optimization - 21st International Conference, IPCO 2020*, pages 91–103, 2020. doi:10.1007/978-3-030-45771-6\_8.
- 6 Jarosław Byrka and Karen Aardal. An optimal bifactor approximation algorithm for the metric uncapacitated facility location problem. *SIAM J. Comput.*, 39(6):2212–2231, 2010. doi:10.1137/070708901.
- 7 Jarosław Byrka, Thomas W. Pensyl, Bartosz Rybicki, Aravind Srinivasan, and Khoa Trinh. An improved approximation for  $k$ -median and positive correlation in budgeted optimization. *ACM Trans. Algorithms*, 13(2):23:1–23:31, 2017. doi:10.1145/2981561.
- 8 Moses Charikar and Sudipto Guha. Improved combinatorial algorithms for facility location problems. *SIAM J. Comput.*, 34(4):803–824, 2005. doi:10.1137/S0097539701398594.

---

<sup>4</sup> Throughout this paper, by guessing we mean trying all the (polynomially many) possible options. Each such option leads to a different solution, and we return the best one.

- 9 Chandra Chekuri and Alina Ene. Approximation algorithms for submodular multiway partition. In *IEEE 52nd Annual Symposium on Foundations of Computer Science, FOCS 2011*, pages 807–816, 2011. doi:10.1109/FOCS.2011.34.
- 10 Chandra Chekuri and Alina Ene. Submodular cost allocation problem and applications. In *Automata, Languages and Programming - 38th International Colloquium, ICALP 2011*, pages 354–366, 2011. doi:10.1007/978-3-642-22006-7\_30.
- 11 Vincent Cohen-Addad, Hossein Esfandiari, Vahab S. Mirrokni, and Shyam Narayanan. Improved approximations for Euclidean  $k$ -means and  $k$ -median, via nested quasi-independent sets. In *54th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2022*, pages 1621–1628, 2022. doi:10.1145/3519935.3520011.
- 12 Vincent Cohen-Addad, Fabrizio Grandoni, Euiwoong Lee, and Chris Schwiegelshohn. Breaching the 2 LMP approximation barrier for facility location with applications to  $k$ -median. In *34th ACM-SIAM Symposium on Discrete Algorithms, SODA 2023*, pages 940–986, 2023. doi:10.1137/1.9781611977554.ch37.
- 13 Alina Ene, Jan Vondrák, and Yi Wu. Local distribution and the symmetry gap: Approximability of multiway partitioning problems. In *24th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2013*, pages 306–325, 2013. doi:10.1137/1.9781611973105.23.
- 14 Jittat Fakcharoenphol, Satish Rao, and Kunal Talwar. A tight bound on approximating arbitrary metrics by tree metrics. *J. Comput. Syst. Sci.*, 69(3):485–497, 2004. doi:10.1016/J.JCSS.2004.04.011.
- 15 Fedor V. Fomin, Dieter Kratsch, and Gerhard J. Woeginger. Exact (exponential) algorithms for the dominating set problem. In *Graph-Theoretic Concepts in Computer Science, 30th International Workshop, WG 2004*, pages 245–256, 2004. doi:10.1007/978-3-540-30559-0\_21.
- 16 Satoru Fujishige. *Submodular functions and optimization*, volume 58 of *Annals of Discrete Mathematics*. Elsevier, 2nd edition, 2005.
- 17 Naveen Garg, Anupam Gupta, Stefano Leonardi, and Piotr Sankowski. Stochastic analyses for online combinatorial optimization problems. In *19th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2008*, pages 942–951, 2008. URL: <http://dl.acm.org/citation.cfm?id=1347082.1347185>.
- 18 Kishen N. Gowda, Thomas W. Pensyl, Aravind Srinivasan, and Khoa Trinh. Improved bi-point rounding algorithms and a golden barrier for  $k$ -median. In *34th ACM-SIAM Symposium on Discrete Algorithms, SODA 2023*, pages 987–1011, 2023. doi:10.1137/1.9781611977554.ch38.
- 19 Fabrizio Grandoni, Anupam Gupta, Stefano Leonardi, Pauli Miettinen, Piotr Sankowski, and Mohit Singh. Set covering with our eyes closed. *SIAM J. Comput.*, 42(3):808–830, 2013. doi:10.1137/100802888.
- 20 Fabrizio Grandoni, Rafail Ostrovsky, Yuval Rabani, Leonard J. Schulman, and Rakesh Venkat. A refined approximation for Euclidean  $k$ -means. *Inf. Process. Lett.*, 176:106251, 2022. doi:10.1016/j.ipl.2022.106251.
- 21 Sudipto Guha and Samir Khuller. Greedy strikes back: Improved facility location algorithms. *J. Algorithms*, 31(1):228–248, 1999. doi:10.1006/JAGM.1998.0993.
- 22 Anupam Gupta, Martin Pál, R. Ravi, and Amitabh Sinha. Sampling and cost-sharing: Approximation algorithms for stochastic optimization problems. *SIAM J. Comput.*, 40(5):1361–1401, 2011. doi:10.1137/080732250.
- 23 Shalmoli Gupta. *Approximation algorithms for clustering and facility location problems*. PhD thesis, University of Illinois Urbana-Champaign, USA, 2018. URL: <https://hdl.handle.net/2142/102419>.
- 24 Nicole Immorlica, David R. Karger, Maria Minkoff, and Vahab S. Mirrokni. On the costs and benefits of procrastination: Approximation algorithms for stochastic combinatorial optimization problems. In *15th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2004*, pages 691–700, 2004. URL: <http://dl.acm.org/citation.cfm?id=982792.982898>.

- 25 Satoru Iwata, Lisa Fleischer, and Satoru Fujishige. A combinatorial strongly polynomial algorithm for minimizing submodular functions. *Journal of the ACM (JACM)*, 48(4):761–777, 2001. doi:10.1145/502090.502096.
- 26 Kamal Jain, Mohammad Mahdian, Evangelos Markakis, Amin Saberi, and Vijay V. Vazirani. Greedy facility location algorithms analyzed using dual fitting with factor-revealing LP. *Journal of the ACM (JACM)*, 50(6):795–824, 2003. doi:10.1145/950620.950621.
- 27 Kamal Jain and Vijay V. Vazirani. Approximation algorithms for metric facility location and  $k$ -median problems using the primal-dual schema and Lagrangian relaxation. *Journal of the ACM (JACM)*, 48(2):274–296, 2001. doi:10.1145/375827.375845.
- 28 Shi Li. A 1.488 approximation algorithm for the uncapacitated facility location problem. *Inf. Comput.*, 222:45–58, 2013. doi:10.1016/J.IC.2012.01.007.
- 29 Shi Li and Ola Svensson. Approximating  $k$ -median via pseudo-approximation. *SIAM J. Comput.*, 45(2):530–547, 2016. doi:10.1137/130938645.
- 30 Mohammad Mahdian, Yinyu Ye, and Jiawei Zhang. Approximation algorithms for metric facility location problems. *SIAM J. Comput.*, 36(2):411–432, 2006. doi:10.1137/S0097539703435716.
- 31 Adam Meyerson. Online facility location. In *42nd Annual Symposium on Foundations of Computer Science, FOCS 2001*, pages 426–431, 2001. doi:10.1109/SFCS.2001.959917.
- 32 Alexander Schrijver. *Theory of linear and integer programming*. John Wiley & Sons, 1999.
- 33 David B. Shmoys, Chaitanya Swamy, and Retsef Levi. Facility location with service installation costs. In *15th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2004*, pages 1088–1097, 2004. URL: <http://dl.acm.org/citation.cfm?id=982792.982953>.
- 34 David B. Shmoys, Éva Tardos, and Karen Aardal. Approximation algorithms for facility location problems (extended abstract). In *29th Annual ACM Symposium on the Theory of Computing, STOC 1997*, pages 265–274, 1997. doi:10.1145/258533.258600.
- 35 Zoya Svitkina and Éva Tardos. Facility location with hierarchical facility costs. In *Proceedings of the Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2006*, pages 153–161, 2006. URL: <http://dl.acm.org/citation.cfm?id=1109557.1109576>.
- 36 Zoya Svitkina and Éva Tardos. Facility location with hierarchical facility costs. *ACM Trans. Algorithms*, 6(2):37:1–37:22, 2010. doi:10.1145/1721837.1721853.
- 37 Jiawei Zhang, Bo Chen, and Yinyu Ye. A multiexchange local search algorithm for the capacitated facility location problem. *Math. Oper. Res.*, 30(2):389–403, 2005. doi:10.1287/MOOR.1040.0125.

## A

 Some Omitted Proofs about SFL

**Proof of Lemma 1.2.** Considering the dual of (Conf-LP):

$$\max \left\{ \sum_{c \in C} \alpha_c + \sum_{f \in F} \beta_f : \sum_{c \in R} \alpha_c + \beta_f \leq g(R) + \sum_{c \in R} d(c, f), \forall R \subseteq C, \forall f \in F \right\}. \quad (\text{Conf-DLP})$$

Notice that for fixed  $\alpha$  and  $\beta$ , the functions  $g_f(R) := g(R) + \sum_{c \in R} d(c, f) - \sum_{c \in R} \alpha_c - \beta_f$  are submodular. Thus, a call of a separation oracle on (Conf-DLP) is equivalent to a minimization of all functions  $g_f(\cdot)$ , which can be done using polynomially many oracle calls of  $g(\cdot)$  [25]. Therefore, an optimal primal solution with  $\text{poly}(N)$  many non-zero variables for (Conf-LP) can be found in polynomial time [32, Corollary 14.1g(v)]. ◀

**Proof of Lemma 1.7.** Let  $I = (C, F, d, g(\cdot))$  be the considered instance of SFL. Consider the complete weighted graph on nodes  $C \cup F$ , with weights induced by  $d$ . For each client  $c$ , let  $f(c)$  be the facility closest to  $c$ . We create a dummy facility  $f'(c)$  and add a dummy edge  $\{c, f'(c)\}$  of weight  $d(c, f(c))$ . Let  $F'$  be the set of newly created facilities. Observe



that  $|F'| = n$ . Finally we remove  $F$  and consider the metric  $d'$  over  $C \cup F'$  induced by the distances over the resulting graph. Let  $I' = (C, F', d', g(\cdot))$  be the obtained instance of SFL. Given a solution  $\varphi'$  for  $I'$ , we obtain a solution  $\varphi$  for  $I$  by simply assigning to  $f(c)$  each client  $c'$  assigned to  $f'(c)$  in  $\varphi'$ .

Let us analyze the approximation factor introduced by this reduction. We first observe that  $\text{cost}_I(\varphi) \leq \text{cost}_{I'}(\varphi')$ . Indeed,  $\text{open}_I(\varphi) = \text{open}_{I'}(\varphi')$ . Furthermore, for each client  $c'$  assigned to  $f'(c)$  by  $\varphi'$ , the associated connection cost w.r.t.  $I$  is  $d(c', f(c)) \leq d(c', c) + d(c, f(c)) = d'(c', f'(c))$ . Hence  $\text{conn}_I(\varphi) \leq \text{conn}_{I'}(\varphi')$ .

Next consider an optimal solution  $\text{opt}$  for  $I$ . For each facility  $f$  with  $\text{opt}^{-1}(f) \neq \emptyset$ , let  $c \in \text{opt}^{-1}(f)$  be the client closest to  $f$ . We define a solution  $\text{opt}'$  for  $I'$  by assigning all the clients in  $\text{opt}^{-1}(f)$  to  $f'(c)$ . Again,  $\text{open}_I(\varphi) = \text{open}_{I'}(\varphi')$ . For each client  $c'$  assigned to  $f$  in  $\text{opt}$ , its connection cost in  $I'$  is

$$d'(c', f'(c)) = d(c, c') + d(c, f(c)) \leq d(c', f) + d(c, f) + d(c, f(c)) \leq d(c', f) + 2d(c, f) \leq 3d(c', f).$$

Hence  $\text{conn}_{I'}(\text{opt}') \leq 3 \text{conn}_I(\text{opt})$ . The claim follows.  $\blacktriangleleft$

**Proof of Lemma 1.8.** Let us guess the value  $L = \max_{c \in C} d(c, \text{opt}(c))$  for some optimal solution  $\text{opt}$ . W.l.o.g. assume  $L > 0$ , otherwise the problem is trivial. Consider the complete weighted graph on nodes  $C \cup F$  with weights induced by  $d$ . Remove the edges of weight larger than  $L$ . We next compute a feasible solution in each connected component of the resulting graph separately. Notice that this part of the reduction is approximation preserving since no client can be assigned to a facility in a different connected component in  $\text{opt}$ .

Let  $C'$  and  $F'$  be the clients and facilities, resp., in one such connected component  $G'$ ,  $X' = C' \cup F'$ , and  $d'$  be the metric induced by the distances in  $G'$ . Consider the corresponding SFL instance  $I' = (C', F', d', g(\cdot))$ . Notice that in each such instance  $I'$  one has  $d'_{\max} \leq NL$ . We next change the location of elements of  $X'$  as follows. We consider the ball  $B(x) := \{y \in X' : d'(x, y) \leq \frac{\varepsilon}{2n}L\}$  of radius  $\frac{\varepsilon}{2n}L$  around each  $x \in X'$ . Let  $\mathcal{I}$  be a maximal (independent) set of such balls so that, if  $B(x), B(y) \in \mathcal{I}$  for  $x \neq y$ , then  $B(x) \cap B(y) = \emptyset$ . For each  $y$  with  $B(y) \notin \mathcal{I}$ , we consider any  $B(x) \in \mathcal{I}$  with  $B(x) \cap B(y) \neq \emptyset$  (which must exist since  $\mathcal{I}$  is maximal) and colocate  $y$  with  $x$ . Let  $I'' = (C', F', d'', g(\cdot))$  be the resulting instance of SFL. Observe that  $d''_{\max} \leq NL$  and  $d''_{\min} \geq \frac{\varepsilon}{n}L$ .

Let  $\tilde{I}$  be the union of all the instances  $I''$ , and  $\tilde{d}$  be the associated distances (where inter-component distances can be considered to be  $+\infty$ ). Given a solution  $\varphi$  for  $\tilde{I}$  (obtained by the union of all the solutions obtained for each instance  $I''$ ), we return exactly the same solution  $\varphi$  for  $I$ .

Let us analyze the approximation factor. Notice that  $\text{open}_I(\varphi) = \text{open}_{\tilde{I}}(\varphi)$ . Furthermore, for each client  $c$ ,  $d(c, \varphi(c)) \leq \tilde{d}(c, \varphi(c)) + \frac{2\varepsilon}{n}L$ , where in the latter term we consider the fact that each client and facility is moved at most at distance  $\frac{\varepsilon}{n}L$  from the original location. Hence  $\text{conn}_I(\varphi) \leq \text{conn}_{\tilde{I}}(\varphi) + 2\varepsilon L$ . Given an optimum solution  $\text{opt}$  for  $I$ , by a symmetric argument one has  $\text{cost}_{\tilde{I}}(\text{opt}) \leq \text{cost}_I(\text{opt}) + 2\varepsilon L \leq (1 + 2\varepsilon) \text{cost}_I(\text{opt})$ , where we used the fact that  $\text{cost}_I(\text{opt}) \geq L$ . Altogether an  $\alpha \geq 1$  approximation algorithm for each instance  $I''$  implies an  $\alpha(1 + 2\varepsilon) + 2\varepsilon \leq \alpha(1 + 4\varepsilon)$  approximation for  $I$ .

Finally, we scale the distance  $d''$  and  $g(\cdot)$  by the same factor  $\frac{2n}{\varepsilon L}$  so that  $d''_{\min} = 2$  and  $d''_{\max} \leq \frac{2nN}{\varepsilon}$ . Clearly this final scaling is approximation preserving.  $\blacktriangleleft$

## B Generalizations of SFL

In this section we discuss some generalizations of SFL.

### B.1 Reduction of the Number of Facilities

In this section we consider the generalization of SFL, next called AFFINE SFL, where the opening cost of each facility  $f$  with assigned clients  $R \neq \emptyset$  is  $g_f(R) := p_f + w_f \cdot g(R)$ , where  $p_f, w_f \geq 0$  are input values. Notice that this generalizes SFL WITH ADDITIVE (resp., MULTIPLICATIVE) OPENING COSTS. We also observe that each  $g_f(\cdot)$  is non-negative monotone submodular.

We show how to reduce to the case where  $m = \text{poly}(n)$  (hence  $N = \text{poly}(n)$ ) while loosing a constant factor in the approximation. We will use this reduction in the following sections to convert an  $O(\log \log N)$  approximation into an  $O(\log \log n)$  one.

► **Lemma B.1.** *For any constant  $\varepsilon > 0$ , there is a  $(3 + 37\varepsilon)$ -approximate reduction from AFFINE SFL to the special case where the number of facilities is  $O_\varepsilon(n^3)$ .*

**Proof.** First of all, consider the case  $m \geq 2^n$ . In this case we can solve the problem optimally in polynomial time via the following reduction to the WEIGHTED SET COVER problem. For an instance  $I = (C, F, d, g(\cdot))$  of AFFINE SFL, consider the instance  $J = (\mathcal{U}, \mathcal{R}, \kappa)$  of WEIGHTED SET COVER with universe  $\mathcal{U} = C$ , set collection  $\mathcal{R} = 2^C$  and weight function  $\kappa$  given as  $\kappa_R = 0$  if  $R = \emptyset$  and  $\kappa_R = \min_{f \in F} (p_f + w_f \cdot g(R) + \sum_{c \in R} d(c, f))$  for  $R \in 2^C \setminus \{\emptyset\}$  (which can be computed in  $\text{poly}(N)$  time). Notice that  $2^{|\mathcal{U}|} = 2^n$  which is polynomially bounded in the input size of  $I$ . The optimal solution to  $J$  induces a solution of exactly the same cost to  $I$  and vice versa. There is a simple dynamic program which solves WEIGHTED SET COVER in time  $O(2^{|\mathcal{U}|} \cdot |\mathcal{U}| \cdot |\mathcal{R}|)$  [15, Lemma 2]. Applying this algorithm to  $J$ , one obtains an optimal solution for the input instance  $I$  in time  $O(2^n \cdot \text{poly}(n, m))$ , which is polynomial in  $m$ .

Hence it remains to consider the case  $m \leq 2^n$ . We show how to reduce the number of facilities to  $O_\varepsilon(n^2 \log(nN)) = O_\varepsilon(n^3)$ , while losing the approximation factor in the claim. By exactly the same reduction as in Lemma 1.8, we can assume that in the input metric  $d$  the maximum distance is  $0 < d_{\max} \leq NL$  and the minimum non-zero distance is  $d_{\min} \geq \frac{\varepsilon}{n}L$  while loosing a factor  $(1 + 4\varepsilon)$  in the approximation. Here  $L$  is some value that lower bounds the cost of a given optimum solution  $\text{opt}$ . Let us guess the largest value  $P$  of  $p_f$  over the facilities with at least one assigned client in  $\text{opt}$ . We discard all the facilities  $f$  with  $p_f > P$ . Now, assuming  $P > 0$ , we replace each  $p_f$  with the value  $p'_f := \lceil \frac{p_f n}{\varepsilon P} \rceil \cdot \frac{\varepsilon P}{n}$  ( $p'_f = p_f$  for  $P = 0$ ). Notice that this can only increase the cost of a given solution  $\varphi$ , however this increase is upper bounded by  $n \cdot \frac{\varepsilon P}{n} \leq \varepsilon \cdot \text{cost}_I(\text{opt})$ , where  $I$  is the input instance of the problem. Hence this reduction preserves the approximation guarantee up to a factor  $1 + \varepsilon$ . After this reduction, the set  $\mathcal{P}'$  of different possible values of  $p'_f$  has cardinality at most  $\frac{n}{\varepsilon}$ .

Let  $I = (C, F, d, p', w, g(\cdot))$  be the instance of AFFINE SFL obtained after the above two reductions. Consider the complete edge-weighted graph on nodes  $C \cup F$ , with weights induced by  $d$ . We modify this graph as follows. For each client  $c$  and value  $p' \in \mathcal{P}'$ , we consider the set of facilities  $F_{p'}$  with  $p'_f = p'$ . Let  $F_{p'}(c, i)$ ,  $i \geq 0$ , be the facilities in  $F_{p'}$  whose distances from  $c$  are in the range  $[\frac{\varepsilon}{n}L \cdot (1 + \varepsilon)^i, \frac{\varepsilon}{n}L \cdot (1 + \varepsilon)^{i+1})$ . We also define the set  $F_{p'}(c, -1)$  of the facilities in  $F_{p'}$  at distance 0 from  $c$ . Notice that there are at most  $1 + \lceil \log_{1+\varepsilon} \frac{nN}{\varepsilon} \rceil$  sets  $F_{p'}(c, i)$  which are non-empty. For each  $F_{p'}(c, i) \neq \emptyset$ , we choose a facility  $f = f_{p'}(c, i)$  with minimum value of  $w_f$ . We create a dummy facility  $f' = f'_{p'}(c, i)$  with opening cost  $g'_{f'}(C') = p' + w_f \cdot g(C')$  for  $C' \neq \emptyset$ , and add a dummy edge  $\{c, f'\}$  of weight  $d(c, f)$ . Let  $F'$

be the set of dummy facilities. Notice that, considering also the previous reduction, one has  $|F'| \leq n \cdot \frac{n}{\varepsilon} \cdot (1 + \lceil \log_{1+\varepsilon} \frac{nN}{\varepsilon} \rceil) = O(n^2 \log(nN))$ . We remove the original facilities  $F$ , and let  $d'$  be the metric given by the distances in the resulting graph  $G'$  on nodes  $C \cup F'$ . We solve the problem on the resulting instance  $I' = (C, F', d', p', w, g(\cdot))$ . Given a solution  $\varphi'$  for  $I'$ , we obtain a solution  $\varphi$  for  $I$  naturally as follows: if  $\varphi'(c') = f'_{p'}(c, i)$ , we assign  $c'$  to  $f_{p'}(c, i)$ .

Let us analyze the approximation factor of this final reduction. The opening costs of  $\varphi$  and  $\varphi'$  are identical. Furthermore, for each client  $c'$  assigned to  $f = f_{p'}(c, i)$  in  $\varphi$ , and for  $f' = f'_{p'}(c, i)$ , one has  $d(c', f) \leq d(c', c) + d(c, f) = d'(c', c) + d'(c, f') = d'(c', f')$ . Hence  $\text{cost}_I(\varphi) = \text{cost}_{I'}(\varphi')$ .

Next consider an optimum solution  $\text{opt}$  for  $I$ . We construct a feasible solution  $\text{opt}'$  for  $I'$  as follows. Let  $S^f \neq \emptyset$  be the clients assigned to some  $f \in F$  in  $\text{opt}$ . Recall that the opening cost of  $f$  is  $g'_f(S^f) = p'_f + w_f \cdot g(S^f)$ . Let  $c \in S^f$  be the client at minimum distance  $d(c, f)$  from  $f$ . Define  $i$  as  $-1$  if  $d(c, f) = 0$ , and otherwise,  $i$  such that  $d(c, f) \in [\frac{\varepsilon}{n}L \cdot (1+\varepsilon)^i, \frac{\varepsilon}{n}L \cdot (1+\varepsilon)^{i+1})$ . In  $\text{opt}'$  we reassign all the clients in  $S^f$  to  $f' = f'_{p'_f}(c, i)$ . The opening cost associated with  $f'$  in  $\text{opt}'$  is no larger than the corresponding cost in  $\text{opt}$  since

$$p'_{f'} + w_{f'} \cdot g(S^{f'}) = p'_f + w_{f'} \cdot g(S^f) \leq p'_f + w_f \cdot g(S^f).$$

In the last inequality above we used the fact that  $f \in F_{p'_f}(c, i)$  and  $f_{p'_f}(c, i)$  is the facility in the latter set with minimum  $w_f$  value. The connection cost of each  $c' \in S^f$  w.r.t.  $\text{opt}'$  satisfies

$$\begin{aligned} d'(c', f') &= d'(c', c) + d'(c, f') = d(c, c') + d(c, f_{p'_f}(c, i)) \\ &\leq d(c', f) + d(c, f) + (1 + \varepsilon)d(c, f) \leq (3 + \varepsilon)d(c', f). \end{aligned}$$

Altogether,  $\text{cost}_{I'}(\text{opt}') \leq (3 + \varepsilon) \text{cost}_I(\text{opt})$ . Considering also the first two reductions, we obtain a global reduction which preserves the approximation guarantee up to a factor  $(1 + 4\varepsilon)(1 + \varepsilon)(3 + \varepsilon) \leq 3 + 37\varepsilon$ .  $\blacktriangleleft$

## B.2 SFL with Multiplicative Opening Costs

In this section we sketch the proof of Theorem 1.3. By Lemma B.1, it is sufficient to provide an  $O(\log \log N)$  approximation.

For  $f \in F$  and  $R \subseteq C$  let  $g_f(R) := w_f \cdot g(R)$ . Note that  $g_f(\cdot)$  is submodular, monotone and has  $g_f(\emptyset) = 0$  for every  $f \in F$ . For any (partial) assignment  $S = (S^f)$  and any vector  $(x_R^f)_{R \subseteq C}^{f \in F}$  let also  $\text{open}'(S) := \sum_{f \in F} g_f(S^f)$ , resp.  $\text{open}'(x) := \sum_{f \in F} \sum_{R \subseteq C} g_f(R) \cdot x_R^f$  and  $\text{cost}'(S) := \text{open}'(S) + \text{conn}(S)$  resp.  $\text{cost}'(x) := \text{open}'(x) + \text{conn}(x)$ .

By these definitions, the LP-relaxation of the MULTSFL is given by the constraints from (Conf-LP) and the objective  $\text{cost}'(\cdot)$ . In particular, the LP-relaxation of MULTSFL can be solved with the approach from Lemma 1.2. We keep the merging rule defined in Section 1.4 and the sampling procedure from Section 2. It is easy to verify that the vector  $\tilde{x}$  resulting from this procedure fulfills Lemma 2.2 w.r.t.  $\text{open}'$  instead of  $\text{open}$ .

We reduce MULTSFL to a similar problem to DLA which we call DLA\* which is the same problem as DLA and with the same input variables as DLA, additional inputs  $\tilde{w}_f \geq 0$  for every  $f \in \tilde{F}$  and cost  $\text{cost}_{\text{DLA}^*}^*(\varphi) = \sum_{f \in \tilde{F}} h_f(\varphi^{-1}(f))$  where  $h_f(\cdot) := \tilde{w}_f h(\cdot)$  for every  $f \in \tilde{F}$ . Its convex relaxation is given by the constraints in (DLA-CP) with the cost function  $\text{cost}_{\text{DLA}^*}^*(z) := \sum_{f \in \tilde{F}} \hat{h}_f(z^f)$  (where  $\hat{h}_f$  is the Lovász extension of  $h_f$ ). The reduction described in Lemma 3.2 can be reproduced to reduce MULTSFL to DLA\*. We define the input values

of DLA\* w.r.t. MULTSFL in the same way we define the input values of DLA w.r.t. SFL, with additionally  $\tilde{w}_f = w_f$  for every  $f \in F$ . Notice that  $h_f(\cdot) = \tilde{w}_f h(\cdot) = g_f(\cdot) = w_f g(\cdot)$ . Every reasoning made in the proof of Lemma 3.2 stays valid.

We now adjust Algorithm 1 for DLA\* as follows: in Step 3, we select the facility  $f_v \in \tilde{F}_v$  with minimum weight  $\tilde{w}_{f_v}$ . In the if-clause 4, we search and verify for supportedness w.r.t.  $h_{f_v}$  instead of  $h$  (which is equivalent unless  $\tilde{w}_{f_v} = 0$ , in which case  $L_\theta(z^{f_v})$  is supported for every  $\theta$ ). Since the new algorithm functions exactly like Algorithm 1, except for an arbitrary selection step becoming determined (in particular, the new algorithm is a possible implementation of Algorithm 1), its correctness is implied by the correctness of Algorithm 1.

Notice that since  $f_v$  in Step 3 is now chosen to have minimal weight, we have for any  $f' \in \tilde{F}_v \setminus \{f_v\}$

$$\hat{h}_{f_v}(z^{f_v} + z^{f'}) \leq \hat{h}_{f_v}(z^{f_v}) + \hat{h}_{f_v}(z^{f'}) \leq \hat{h}_{f_v}(z^{f_v}) + \hat{h}_{f'}(z^{f'}),$$

which means that the cost of  $z$  does not increase at any time by the arguments as before. Also, notice that since  $h_f$  is submodular, monotone and  $h_f(\emptyset) = 0$  we can apply Lemma 3.4 with respect to  $h_{f_v}$  instead of  $h$ . Thus, the cost of the sets added at Step 5 and Step 7 is still bounded as in (4) and (5).

### B.3 SFL with Additive Opening Costs

In this section we sketch the proof of Theorem 1.4. As in the previous section, by Lemma B.1, it is sufficient to provide an  $O(\log \log N)$  approximation.

Similarly to the previous section, we define the set function  $g_f(\cdot)$  as  $g_f(R) = g(R) + p_f$  for  $R \neq \emptyset$  and  $g_f(\emptyset) = 0$ . As argued in the previous section, we can find an optimum to the LP relaxation of ADDSFL and reduce it to the problem DLA\* as defined in the last section, but with input weights  $\tilde{p}_f$  instead of  $\tilde{w}_f$  and  $h_f(\cdot)$  as  $h_f(R) := h(R) + p_f$  for  $R \neq \emptyset$ , and  $h_f(\emptyset) = 0$ .

We adapt Algorithm 1 like in the previous section: in Step 3, we select the facility  $f_v \in \tilde{F}_v$  with minimum weight  $\tilde{p}_{f_v}$ . In the if-clause 4, we search and verify for supportedness w.r.t.  $h_{f_v}$  instead of  $h$ . The correctness of the new algorithm here is given by the same argument as in the previous section. Notice that by (2) we have  $\hat{h}_f(z) = \hat{h}(z) + p_f \cdot \max_{c \in \tilde{C}} z_c$ , which implies  $\hat{h}_{f_v}(z^{f_v} + z^{f'}) \leq \hat{h}_{f_v}(z^{f_v}) + \hat{h}_{f'}(z^{f'})$  with  $f_v$  chosen as in Step 3 in Algorithm 1. The cost of  $z$  does therefore not increase throughout the algorithm. Bounding the cost of sets added to the solution at Step 5 and Step 7 can be done, like for MULTSFL, by applying Lemma 3.4 to  $h_{f_v}$ .