Strategyproof Facility Location for Concave Cost Functions[*]

Dimitris Fotakis[1] and Christos Tzamos[2]

[1] School of Electrical and Computer Engineering,
National Technical University of Athens, 157 80 Athens, Greece.
Email: `fotakis@cs.ntua.gr`
[2] Computer Science and Artificial Intelligence Laboratory,
Massachusetts Institute of Technology, Cambridge, MA 02139.
Email: `tzamos@mit.edu`

Abstract. We consider k-Facility Location games, where n strategic agents report their locations on the real line and a mechanism maps them to k facilities. Each agent seeks to minimize his connection cost, given by a nonnegative increasing function of his distance to the nearest facility. Departing from previous work, that mostly considers the identity cost function, we are interested in mechanisms without payments that are (group) strategyproof for any given cost function, and achieve a good approximation ratio for the social cost and/or the maximum cost of the agents.

We present a randomized mechanism, called EQUAL COST, which is group strategyproof and achieves a bounded approximation ratio for all k and n, for any given concave cost function. The approximation ratio is at most 2 for MAX COST and at most n for SOCIAL COST. To the best of our knowledge, this is the first mechanism with a bounded approximation ratio for instances with $k \geq 3$ facilities and any number of agents. Our result implies an interesting separation between deterministic mechanisms, whose approximation ratio for MAX COST jumps from 2 to unbounded when k increases from 2 to 3, and randomized mechanisms, whose approximation ratio remains at most 2 for all k. On the negative side, we exclude the possibility of a mechanism with the properties of EQUAL COST for strictly convex cost functions. We also present a randomized mechanism, called PICK THE LOSER, which applies to instances with k facilities and only $n = k + 1$ agents. For any given concave cost function, PICK THE LOSER is strongly group strategyproof and achieves an approximation ratio of 2 for SOCIAL COST.

Keywords: Algorithmic Mechanism Design; Social Choice; Facility Location Games

1 Introduction

We consider k-*Facility Location games*, where k facilities are placed on the real line based on the preferences of n strategic agents. Such problems are motivated by natural scenarios in Social Choice, where the government plans to build a fixed number of public facilities in an area (see e.g., [13]). The choice of the locations is based on the preferences of local people, or *agents*. Each agent reports his ideal location, and the government applies a (deterministic or randomized) *mechanism* that maps the agents' preferences to k facility locations.

The agents evaluate the outcome of the mechanism according to their *connection cost*, given by a nonnegative increasing function $c(d)$ of the distance d of their ideal location to the nearest facility. Agents seek to minimize their connection cost, and may misreport their ideal locations in an attempt of manipulating the mechanism. Therefore, the mechanism should be *strategyproof*, i.e., should ensure that no agent can benefit from misreporting his location, or even *group strategyproof*, i.e., should ensure that for any coalition of agents misreporting their locations, at least one of them does not benefit. The government's goal is to minimize an objective function of the agents' connection cost. Most prominent among them are the objective of SOCIAL COST, which considers the total cost of the agents, and the objective of MAX COST, which considers the maximum cost of an agent. So, in addition to (group) strategyproofness, the mechanism should either optimize or achieve a reasonable approximation to the designated objective function, thus ensuring that the outcome is socially efficient.

Previous Work. The numerous applications and the elegance of the model have attracted a significant volume of research on the problem. In Social Choice, the emphasis has been on characterizing the class of (group) strategyproof mechanisms for locating a single facility if the agents' preferences are *single-peaked*. Roughly speaking, an agent has single-peaked preferences if he has an ideal location (or *peak*), and consistently prefers less the locations farther from it. However, the strength of his preference for locations closer to his peak is not explicitly quantified by any function of the distance. For general single-peaked preferences, a classical result of Moulin [14] shows that the class of deterministic strategyproof mechanisms for locating a single facility on the line coincides with the class of generalized median mechanisms (see also the surveys of Barberá [3] and Sprumont [19], and [15, Chapter 10]). Schummer and Vohra [18] extended this characterization to tree metrics, and proved that for non-tree metrics, any onto strategyproof mechanism must be a dictatorship. More recently, Dokow et al. [4] obtained similar characterizations for locating a single facility on the discrete line and on the discrete circle.

Adopting an optimization viewpoint to Facility Location games, Procaccia and Tennenholtz [17] introduced the framework of *approximate mechanism design without money*. The basic idea is to consider game-theoretic versions of optimization problems, such as k-Facility Location, where efficiency is quantified by an objective function (instead of efficiency related properties, such as onto, non-dictatorship, and Pareto-efficiency, typically studied in Social Choice). Then, any reasonable approximation to the optimal solution can be regarded as a socially desirable outcome, and one seeks to determine the best approximation ratio achievable by strategyproof mechanisms. As for the preferences of the agents, with respect to which strategyproofness is defined, this line of research adopted the standard definition of Facility Location problems from Operations Research (see e.g., [12]). Thus, it implicitly abandoned the setting of general single-peaked preferences, in favor of the more restricted (and technically easier to handle) case where the agents' cost is given by a linear function $c(d) = \alpha d$ of their distance d to the nearest facility. Translated into this framework, the results of [14, 18] imply a deterministic strategyproof mechanism that minimizes the SOCIAL COST for 1-Facility Location

MAX COST

	$k = 1$	$k = 2$	$2 < k < n - 1$	$k = n - 1$
Deterministic	2 [17]	2 [17]	∞ [7]	∞ [7]
Randomized	1.5 [17]	$[1.5, 5/3]$ [17]	$[1.5, \mathbf{2}]$ **[here]**	1.5 [5]

SOCIAL COST

	$k = 1$	$k = 2$	$2 < k < n - 1$	$k = n - 1$
Deterministic	1 [14]	$n - 2$ [7], [17]	∞ [7]	∞ [7]
Randomized	1 [14]	$[1.045, 4]$ [11], [10]	$[1.045, \mathbf{n}]$ **[here]**	$[1.045, \mathbf{2}]$ **[here]**

Fig. 1. Summary of known results on the approximability of k-Facility Location on the line (with linear cost functions) for the objectives of MAX COST and SOCIAL COST. In each cell, we have either the precise approximation ratio (if known) or the interval determined by the best known lower and upper bounds. In cells with two references, the first is for the lower bound and the second for the upper bound. We note that the lower bound on the approximation ratio of deterministic mechanisms for $k \geq 3$ is only shown for anonymous mechanisms. The randomized upper bounds proven in this work are shown in bold and hold for any concave cost function.

on the line and in tree metrics. On the negative side, the impossibility result of [18] implies that the best approximation ratio achievable for the objective of SOCIAL COST by deterministic strategyproof mechanisms for 1-Facility Location in general metrics is $n - 1$. However, the explicit quantification of agents' preferences now allows for randomized mechanisms that are strategyproof with respect to the agents' expected cost (a.k.a. incentive compatible in expectation, see e.g., [15, Section 9.5.6]) and may achieve better approximation ratios.

Since [17], there has been a considerable interest in quantifying the best approximation ratio achievable by strategyproof mechanisms for k-Facility Location on the line and in general metric spaces. As a result, the approximability of k-Facility Location (with linear cost functions) by deterministic and randomized strategyproof mechanisms has become well understood in many interesting cases (see also Fig. 1). The main message is that deterministic strategyproof mechanisms can only achieve a bounded approximation ratio if we have at most 2 facilities [17, 7]. On the other hand, randomized mechanisms are known to achieve better approximation ratios for 2-Facility Location and also a bounded approximation ratio if we have $k \geq 2$ facilities and only $n = k+1$ agents [5]. Notably, such instances are known to be hard for deterministic mechanisms. In particular, the inapproximability of k-Facility Location by anonymous deterministic strategyproof mechanisms, for all $k \geq 3$, was proved in [7] for instances with only $n = k + 1$ agents (see also [5] for some concrete examples that motivate the special case where all agents but one get a facility).

Motivation and Contribution. Our work is motivated by two natural questions related to approximate mechanism design without money for k-Facility Location. The first question is about the approximability of k-Facility Location by randomized strategyproof mechanisms for instances with any number of facilities and any number of agents. Prior to this work, we have only known randomized mechanisms with a bounded[3] approximation ratio if we have either at most 3 facilities or k facilities and only $n = k + 1$ agents. Most importantly, all the randomized upper bounds in Fig. 1 are achieved by mechanisms that balance strategyproofness against efficiency using different approaches (see e.g., [17, 10, 5]).

---

[3] The approximation ratio of a mechanism for k-Facility Location is bounded if it is a function of n and k. We highlight that this property is essentially objective-independent, since any mechanism with a bounded approximation ratio for e.g., MAX COST also has a bounded approximation for SOCIAL COST and for the objective of minimizing the L_p norm of the agents' costs, for any $p \geq 1$, and vice versa.

2

The second question is whether the restriction to linear cost functions is a necessary price to pay for adopting the elegant optimization framework of Procaccia and Tennenholtz [17] and aiming at a reasonable approximation ratio. In fact, we can imagine a few natural scenarios where the agents' cost is best described by a convex or a concave non-decreasing cost function $c(d)$ of their distance d to the nearest facility. For example, a convex cost function captures the fact that the growth rate of the people's disutility from commuting increases with the distance (e.g., in addition to cost and time considerations, people get more and more tired if they commute over long distances). On the other hand, a concave cost function captures the fact that the growth rate of the traveling time decreases with the distance (e.g., people walk over short distances, bike over medium distances, drive over long distances, and take a plane over really long ones) and the fact that the traveling cost per unit of distance decreases with the distance (if we take the cost of food into account for shorter distances). To a certain extent, a setting where the agents' cost function is not fixed, but is given as part of the input, would be closer to the setting of general single-peaked preferences in Social Choice. Then, a mechanism should be strategyproof, or even group strategyproof, for any given cost function c, just as generalized median mechanisms are strategyproof for any collection of single-peaked preferences, while the approximation ratio may also depend on some quantitative properties (e.g., the derivative) of c. Notably, this holds for the class of percentile mechanisms [20], which decide on the facility locations based on the ordering of the agents on the line, are group strategyproof, and include the optimal (wrt. the approximation ratio for linear cost functions) deterministic mechanisms for 1 and 2-Facility Location on the line. However, percentile mechanisms have an unbounded approximation ratio for all $k \geq 3$. In contrast, the strategyproofness of known randomized mechanisms crucially depends on the linearity of the cost function (see e.g., [17, Mechanism 1] which is not strategyproof e.g., for $c(d) = \sqrt{d}$).

In this work, we make significant progress in both research directions above. Our main technical contribution consists of two simple and natural randomized mechanisms, called EQUAL COST and PICK THE LOSER, that are group strategyproof and achieve a bounded approximation ratio for any number of facilities and any given concave cost function.

EQUAL COST, presented in Section 3, applies to instances with any number of facilities k and any number of agents n, and is the first (group) strategyproof mechanism with a bounded approximation ratio for all k and n. Its approximation ratio is at most 2 for MAX COST and at most n for SOCIAL COST, for all concave cost functions c. Combined with the lower bound of [17] for the objective of MAX COST, this implies that the best approximation ratio achievable by randomized mechanisms for k-Facility Location on the line is at least 1.5 and at most 2, for all k and for all concave cost functions. Moreover, we obtain an interesting separation between deterministic mechanisms, whose approximation ratio for MAX COST jumps from 2 to unbounded when k increases from 2 to 3, and randomized mechanisms, whose approximation ratio remains a small constant for all k.

From a technical viewpoint, EQUAL COST works by equalizing the expected cost of all agents. The mechanism first covers the agents' locations with k disjoint intervals of length ℓ, where ℓ is chosen so that $c(\ell)$ is at most twice the optimal maximum cost of an agent. Then, taking the cost function c into account, it computes a random variable X in $[0, \ell]$, so that all locations $x \in [0, \ell]$ have the same expected cost, under c, if x is connected to a facility distributed in $[0, \ell]$ according to X. Finally, EQUAL COST places a facility in each interval according to the random variable X so that all agents have an expected cost equal to the expectation of $c(X)$.

The key technical claim in the analysis of EQUAL COST is that if the cost function c is concave and piecewise linear, a random variable X with the desired properties exists and can be computed

3

efficiently as the solution to a homogeneous system of linear equations (Lemma 3.2). This claim can be generalized to any continuous concave function, but the technical details have to do with techniques for the solution of integral equations and are beyond the scope of this work. We show that EQUAL COST is (resp. strongly) group strategyproof for any given (resp. strictly) concave cost function c, and that the agents' expected cost is at most the maximum cost of an agent in the optimal solution for the objective of MAX COST (Lemma 3.5). In addition to implying the approximation guarantees, the upper bound on the expected cost of the agents indicates that the facility allocation of EQUAL COST is fair in expectation, and does not unnecessarily increase the agents' disutility.

To demonstrate the natural behavior of EQUAL COST for typical cost functions, we derive the exact form of the random variable X for three important cases: linear cost functions, piecewise linear cost functions with two pieces, and exponential cost functions of the form $c(d) = 1 - e^{-\lambda d}$ (Section 4). Moreover, we show how to implement EQUAL COST if the agents and the facilities should lie in a bounded interval (Section 5). This implies that EQUAL COST can be applied to instances where the agents lie on a circle metric, with the same approximation guarantees, but rather surprisingly, with group strategyproofness carrying over only if the number of facilities is even.

On the negative side, we exclude the possibility of a mechanism with the properties of EQUAL COST for strictly convex cost functions (Section 5.2). Specifically, we show that the expected cost of the agents in the same interval cannot be equalized if the cost function c is strictly convex. Moreover, employing an exponential cost function, we show (Lemma 5.2) that there does not exist a randomized strategyproof mechanism with a bounded approximation ratio for any given convex cost function (note that the approximation ratio here may also depend on the cost function).

In Section 6, we focus on the simpler and elegant setting where we have k facilities and only $n = k + 1$ agents. This setting was motivated and studied in [5], and deserves special attention not only because such instances are among the hardest ones for deterministic mechanisms (see e.g., [7, Theorem 7.1]), but also because they make EQUAL COST perform poorly for the objective of SOCIAL COST. We present the PICK THE LOSER mechanism that allocates facilities to all but a single agent, designated as the loser. The probability distribution according to which the loser is chosen is motivated by the probability distribution used by [9] for scheduling on selfish unrelated machines. Our key technical contribution here is to show that PICK THE LOSER is strongly group strategyproof for any given concave cost function (Lemma 6.1). We also show that PICK THE LOSER achieves an approximation ratio of 2 for the objective of SOCIAL COST. Thus, we significantly improve on the previously best known approximation ratio of $n/2$ achieved by the INVERSELY PROPORTIONAL mechanism of [5] for this class of instances. Moreover, the small approximation ratio of PICK THE LOSER nicely complements the poor performance of EQUAL COST for such instances.

Other Related Work. For the objective of MAX COST, Alon et al. [1] almost completely characterized the approximation ratios achievable by randomized and deterministic mechanisms for 1-Facility Location in general metrics and rings. For the objective of SOCIAL COST, Nissim et al. [16] and Fotakis and Tzamos [8] considered imposing randomized mechanisms that achieve an additive approximation of $o(n)$ and an approximation ratio of $4k$ for k-Facility Location on the line and in general metric spaces, respectively. In fact, the approximation ratio can be improved to $\Theta(\ln K)$, using the analysis of [2]. For 1-Facility Location on the line and the objective of minimizing the sum of squares of the agents' distances to the facility, Feldman and Wilf [6] proved that the best approximation ratio is 1.5 for randomized and 2 for deterministic mechanisms. Moreover, they presented a class of randomized mechanisms that includes all known strategyproof mechanisms for 1-Facility Location on the line.

2 Notation, Definitions, and Preliminaries

For a random variable X, we let $\mathbb{E}[X]$ denote the *expectation* of X. For an event E in a sample space, we let $\Pr[E]$ denote the probability that E occurs.

Instances. We consider k-Facility Location with $k \geq 1$ facilities and $n \geq k+1$ agents on the real line. We let $N = \{1, \ldots, n\}$ be the set of agents. Each agent $i \in N$ resides at a location $x_i \in \mathbb{R}$, which is i's *private* information (i.e., each location x_i is a piece of information only known by agent i). An *instance* is a tuple $(\boldsymbol{x}, c)$, where $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ is the agents' locations profile and $c : \mathbb{R}_{\geq 0} \mapsto \mathbb{R}_{\geq 0}$ is a cost function that gives the connection cost of each agent. The cost function c is *public knowledge* and the same for all agents. Normalizing c, we assume that $c(0) = 0$. If the cost function c is clear from the context, we let an instance simply consist of $\boldsymbol{x}$.

For an n-tuple $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$, we let $\boldsymbol{x}_{-i} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ be $\boldsymbol{x}$ without x_i. For a non-empty set S of indices, we let $\boldsymbol{x}_S = (x_i)_{i \in S}$ and $\boldsymbol{x}_{-S} = (x_i)_{i \notin S}$. We write $(\boldsymbol{x}_{-i}, a)$ to denote the tuple $\boldsymbol{x}$ with a in place of x_i, $(\boldsymbol{x}_{-\{i,j\}}, a, b)$ to denote the tuple $\boldsymbol{x}$ with a in place of x_i and b in place of x_j, and so on.

Mechanisms. A *deterministic mechanism* F for k-Facility Location maps an instance $(\boldsymbol{x}, c)$ to a k-tuple $(y_1, \ldots, y_k) \in \mathbb{R}^k$, $y_1 \leq \cdots \leq y_k$, of facility locations. We let $F(\boldsymbol{x}, c)$ (or simply $F(\boldsymbol{x})$, whenever c is clear from the context) denote the outcome of F for instance $(\boldsymbol{x}, c)$, and let $F_j(\boldsymbol{x}, c)$ denote y_j, i.e., the j-th smallest coordinate in $F(\boldsymbol{x}, c)$. We write $y \in F(\boldsymbol{x}, c)$ to denote that $F(\boldsymbol{x}, c)$ has a facility at location y. A *randomized mechanism* F maps an instance $(\boldsymbol{x}, c)$ to a probability distribution over k-tuples $(y_1, \ldots, y_k) \in \mathbb{R}^k$.

Connection Cost, Social Cost, Maximum Cost. Given a k-tuple $\boldsymbol{y} = (y_1, \ldots, y_k)$, $y_1 \leq \cdots \leq y_k$, of facility locations, the connection cost of agent i with respect to $\boldsymbol{y}$, denoted $\mathrm{cost}(x_i, \boldsymbol{y})$, is $\mathrm{cost}(x_i, \boldsymbol{y}) = c(\min_{1 \leq j \leq k} |x_i - y_j|)$. Given a deterministic mechanism F and an instance $(\boldsymbol{x}, c)$, we let $\mathrm{cost}(x_i, F(\boldsymbol{x}, c))$ (or simply, $\mathrm{cost}(x_i, F(\boldsymbol{x}))$, if c is clear from the context) denote the connection cost of agent i with respect to the outcome of $F(\boldsymbol{x}, c)$. If F is a randomized mechanism, the expected connection cost of agent i is

$$\mathrm{cost}(x_i, F(\boldsymbol{x}, c)) = \mathbb{E}_{\boldsymbol{y} \sim F(\boldsymbol{x}, c)}[\mathrm{cost}(x_i, \boldsymbol{y})]$$

The MAX COST of a deterministic mechanism F for an instance $(\boldsymbol{x}, c)$ is

$$\mathrm{MC}[F(\boldsymbol{x}, c)] = \max_{i \in N} \mathrm{cost}(x_i, F(\boldsymbol{x}, c))$$

The expected MAX COST of a randomized mechanism F for an instance $(\boldsymbol{x}, c)$ is

$$\mathrm{MC}[F(\boldsymbol{x}, c)] = \mathbb{E}_{\boldsymbol{y} \sim F(\boldsymbol{x}, c)}[\max_{i \in N} \mathrm{cost}(x_i, \boldsymbol{y})]$$

The optimal MAX COST, denoted $\mathrm{MC}^*(\boldsymbol{x}, c)$, is $\mathrm{MC}^*(\boldsymbol{x}, c) = \min_{\boldsymbol{y} \in \mathbb{R}^k} \max_{i \in N} \mathrm{cost}(x_i, \boldsymbol{y})$.

The (resp. expected) SOCIAL COST of a deterministic (resp. randomized) mechanism F for an instance $(\boldsymbol{x}, c)$ is $\mathrm{SC}[F(\boldsymbol{x}, c)] = \sum_{i=1}^{n} \mathrm{cost}(x_i, F(\boldsymbol{x}, c))$. The optimal SOCIAL COST, denoted $\mathrm{SC}^*(\boldsymbol{x}, c)$, is $\mathrm{SC}^*(\boldsymbol{x}, c) = \min_{\boldsymbol{y} \in \mathbb{R}^k} \sum_{i=1}^{n} \mathrm{cost}(x_i, \boldsymbol{y})$.

Approximation Ratio. A (randomized) mechanism F for k-Facility Location achieves an *approximation ratio* of $\rho \geq 1$ for a class of cost functions $\mathcal{C}$ and the objective of MAX COST (resp. SOCIAL COST), if for all cost functions $c \in \mathcal{C}$ and all location profiles $\boldsymbol{x}$, $\mathrm{MC}[F(\boldsymbol{x}, c)] \leq \rho\, \mathrm{MC}^*(\boldsymbol{x}, c)$ (resp. $\mathrm{SC}[F(\boldsymbol{x}, c)] \leq \rho\, \mathrm{SC}^*(\boldsymbol{x}, c)$).

Strategyproofness and Group Strategyproofness. A mechanism F is *strategyproof* for a class of cost functions $\mathcal{C}$ if no agent can benefit from misreporting his location. Formally, F is strategyproof if for all cost functions $c \in \mathcal{C}$, all location profiles $\boldsymbol{x}$, any agent i, and all locations y,

$$\text{cost}(x_i, F(\boldsymbol{x}, c)) \leq \text{cost}(x_i, F((\boldsymbol{x}_{-i}, y), c)).$$

A mechanism F is (weakly) *group strategyproof* for a class of cost functions $\mathcal{C}$ if for any coalition of agents misreporting their locations, at least one of them does not benefit. Formally, F is (weakly) *group strategyproof* if for all cost functions $c \in \mathcal{C}$, all location profiles $\boldsymbol{x}$, any non-empty coalition S, and all location profiles $\boldsymbol{y}_S$ for S, there exists some agent $i \in S$ such that

$$\text{cost}(x_i, F(\boldsymbol{x}, c)) \leq \text{cost}(x_i, F((\boldsymbol{x}_{-S}, \boldsymbol{y}_S), c)).$$

A mechanism F is *strongly group strategyproof* for a class of cost functions $\mathcal{C}$ if there is no coalition S of agents misreporting their locations where at least one agent in S benefits and the other agents in S do not lose from the deviation. Formally, F is strongly group strategyproof if for all cost functions $c \in \mathcal{C}$ and all location profiles $\boldsymbol{x}$, there do not exist a non-empty coalition S and a location profile $\boldsymbol{y}_S$ for S, such that for all $i \in S$,

$$\text{cost}(x_i, F(\boldsymbol{x}, c)) \geq \text{cost}(x_i, F((\boldsymbol{x}_{-S}, \boldsymbol{y}_S), c)),$$

and there exists some agent $j \in S$ with

$$\text{cost}(x_j, F(\boldsymbol{x}, c)) > \text{cost}(x_j, F((\boldsymbol{x}_{-S}, \boldsymbol{y}_S), c)).$$

3 The EQUAL-COST Mechanism

In this section, we present and analyze the EQUAL COST mechanism. At the conceptual level, EQUAL COST, or EC, in short, works by equalizing the expected cost of all agents. Given an instance $(\boldsymbol{x}, c)$ of k-Facility Location on the line, EC works as follows:

Step 1 It computes an optimal covering of all agent locations with k disjoint intervals $[\alpha_i, \alpha_i + \ell]$ that minimizes the interval length ℓ (wlog., we assume that $\alpha_i < \alpha_{i+1}$).
Step 2 It constructs a random variable $X(\ell) \in [0, \ell]$ such that all locations $x \in [0, \ell]$ have the same expected connection cost $\mathbb{E}[c(|x - X|)]$.
Step 3 For every interval $[\alpha_i, \alpha_i + \ell]$, EC places a facility at $\alpha_i + X$, if i is odd, or at $\alpha_i + \ell - X$, if i is even.

We proceed to establish the main properties of EC, summarized by the following theorem. For the proof, we examine, in the following sections, each step of the mechanism separately.

Theorem 3.1. *For the class of all concave cost functions,* EQUAL COST *is group strategyproof and achieves an approximation ratio of* 2 *for the objective of* MAX COST, *and an approximation ratio of* n *for the objective of* SOCIAL COST. *Moreover, for every instance* $(\boldsymbol{x}, c)$, *with c concave, and every agent* i, $\text{cost}(x_i, \text{EC}(\boldsymbol{x}, c)) \leq \text{MC}^*(\boldsymbol{x}, c)$.

3.1 Step 1: Partitioning the Instance in Intervals

We can compute the minimum feasible interval length ℓ by checking all possible candidate values. The value of ℓ is equal to the distance $x_j - x_i$ for some agent locations $x_j \geq x_i$. So, there are at most $n^2/2$ candidate values for ℓ. For each candidate value ℓ', we can check feasibility and compute a covering of all locations in $\boldsymbol{x}$ with intervals of length ℓ' as follows:

> While there are uncovered agents, find the leftmost uncovered agent i, and create a new interval $[x_i, x_i + \ell']$.

The above algorithm computes the minimum number of intervals of length ℓ' to cover $\boldsymbol{x}$. If this number is at most k, we set $\ell = \ell'$. We can also speed up the algorithm by binary search over the space of candidate values.

We observe that the partitioning into intervals of length ℓ is closely related to the optimal maximum cost $\mathrm{MC}^*(\boldsymbol{x}, c)$. In fact, an optimal solution can be obtained by placing a facility at the midpoint of each interval. Thus, the cost of the optimal solution is $\mathrm{MC}^*(\boldsymbol{x}, c) = c(\ell/2)$.

3.2 Step 2: Constructing the Random Variable

We next show that for any given cost function c, we can construct a family of random variables $X(\ell) \in [0, \ell]$ such the expected cost of every point in $[0, \ell]$ is the same. For convenience, we denote this cost as $C(\ell)$. We note that $C(\ell) = \mathbb{E}[c(|X(\ell) - x|)]$, for all $x \in [0, \ell]$. In particular, for $x = 0$, we get $C(\ell) = \mathbb{E}[c(X(\ell))]$.

We assume that the cost function c is piecewise-linear with pieces of length 1 and growth rates $\lambda_0, \lambda_1, \ldots, \lambda_i, \ldots$, where λ_i is the growth rate in the interval $[i, i+1)$. For all i, $\lambda_i > 0$ and $\lambda_i \geq \lambda_{i+1}$, because c is strictly increasing and concave. Our result applies to general concave functions either by discretizing appropriately, or by solving a continuous analog of the homogeneous linear system below through an integral equation. The technical details are related to the solution of integral equations and are beyond the scope of this work.

The support S of the random variable $X(\ell)$ is every point i and $\ell - i$, for integer $i = 0, \ldots, \lfloor \ell \rfloor$. We note that if ℓ is an integer, we have only $|S| = \ell + 1$ points in the support, instead of $|S| = 2(\lfloor \ell \rfloor + 1)$ points in general. The crucial observation is that the derivative of the expected cost function in every interval between consecutive points in the support must be 0. So, to compute the probability p_j assigned to each point j in the support of $X(\ell)$, we write a set of $|S| - 1$ linear equations and $|S|$ unknowns (the probability p_j of each point j in the support) requiring that the derivative of the expected cost function in each interval is 0. So, we get the homogeneous linear system $\Lambda \boldsymbol{p} = \boldsymbol{0}$. If ℓ is an integer, the matrix Λ is:

$$\Lambda = \begin{pmatrix} \lambda_0 & -\lambda_0 & -\lambda_1 & \ldots & -\lambda_{\ell-1} \\ \lambda_1 & \lambda_0 & -\lambda_0 & \ldots & -\lambda_{\ell-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \lambda_{\ell-1} & \lambda_{\ell-2} & \lambda_{\ell-3} & \ldots & -\lambda_0 \end{pmatrix}$$

Namely, the elements of the matrix Λ are $\Lambda_{i,j} = \lambda_{i-j}$, if $i \geq j$, and $\Lambda_{i,j} = -\lambda_{j-i-1}$, if $i < j$, for all $i = 0, \ldots, \ell - 1$ and $j = 0, \ldots, \ell$, where λ_κ denotes the growth rate of the piecewise-linear cost function c at the support point κ.

If ℓ is not an integer, the elements of the matrix Λ are $\Lambda_{i,j} = \lambda_{\lfloor (i-j)/2 \rfloor}$, if $i \geq j$, and $\Lambda_{i,j} = \lambda_{\lfloor (j-i-1)/2 \rfloor}$, if $i < j$, for all $i = 0, \ldots, 2\lfloor \ell \rfloor$ and $j = 0, \ldots, 2\lfloor \ell \rfloor + 1$. Thus,

$$\Lambda = \begin{pmatrix} \lambda_0 & -\lambda_0 & -\lambda_0 & -\lambda_1 & -\lambda_1 & -\lambda_2 & -\lambda_2 & \ldots & -\lambda_{\lfloor \ell \rfloor - 1} & -\lambda_{\lfloor \ell \rfloor - 1} & -\lambda_{\lfloor \ell \rfloor} \\ \lambda_0 & \lambda_0 & -\lambda_0 & -\lambda_0 & -\lambda_1 & -\lambda_1 & -\lambda_2 & -\lambda_2 & \ldots & -\lambda_{\lfloor \ell \rfloor - 1} & -\lambda_{\lfloor \ell \rfloor - 1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \lambda_{\lfloor \ell \rfloor} & \lambda_{\lfloor \ell \rfloor - 1} & \lambda_{\lfloor \ell \rfloor - 1} & \ldots & \ldots & \ldots & \lambda_1 & \lambda_1 & \lambda_0 & \lambda_0 & -\lambda_0 \end{pmatrix}$$

We now show that in both cases there is a unique symmetric probability distribution that satisfies the system of equations. For this purpose, we use the two lemmas below. The first lemma is about a class of diagonally dominant matrices. It shows that we can bring any such matrix into a triangular form by performing Gaussian elimination, such that all diagonal elements are positive and all off-diagonal elements are less than or equal to 0.

Lemma 3.1. *Let A be a $q \times n$, $q \leq n$ matrix so that $A_{i,i} > 0$, for all $i = 1, \ldots, q$, $A_{i,j} \leq 0$, for all $i \neq j$, and $\sum_{i=1}^{q} A_{i,j} > 0$, for all $j = 1, \ldots, q$. Then, by performing elementary row operations (Gaussian elimination) on A, we can get a row-echelon form A' where $A'_{i,i} > 0$, for all $i = 1, \ldots, q$, $A'_{i,j} = 0$, for all $i > j$, and $A'_{i,j} \leq 0$, for all $i < j$.*

Proof. We use induction on q. The base case, where $q = 1$, is already in the desired form. Assuming that the lemma holds for $q \geq 1$, we show that it holds for $q + 1$.

We have that $A = \begin{pmatrix} a & \boldsymbol{u}^T \\ \boldsymbol{v} & B \end{pmatrix}$, with $a > 0$ and all elements of $\boldsymbol{u}$ and $\boldsymbol{v}$ non-positive. With a single step of Gaussian elimination, we get $\begin{pmatrix} a & \boldsymbol{u}^T \\ \boldsymbol{0} & B - \frac{\boldsymbol{v} \times \boldsymbol{u}^T}{a} \end{pmatrix}$. To conclude the induction step, we show that the submatrix $B' = B - \frac{\boldsymbol{v} \times \boldsymbol{u}^T}{a}$ satisfies the properties of the lemma. Since all elements of $\boldsymbol{v} \times \boldsymbol{u}^T$ are non-negative, we still have $B'_{i,j} \leq 0$, for all $i \neq j$. So, we need to show that $\sum_{i=1}^{q} B'_{i,j} > 0$, for all columns $j = 1, \ldots, q$, which also implies that $B'_{i,i} > 0$, for all $i = 1, \ldots, q$. For any column $j = 1, \ldots, q$, we have that:

$$\sum_{i=1}^{q} B'_{i,j} = \sum_{i=1}^{q} (B_{i,j} - v_i u_j / a) = \sum_{i=1}^{q} B_{i,j} - \frac{u_j}{a} \sum_{i=1}^{q} v_i > -u_j - \frac{u_j}{a}(-a) = 0 \,.$$

For the last inequality, we use that $u_j \leq 0$, and the hypothesis that $\sum_{i=1}^{q+1} A_{i,j} > 0$, which implies that $a + \sum_{i=1}^{q} v_i > 0$ and that $u_j + \sum_{i=1}^{q} B_{i,j} > 0$. $\qquad \square$

The next lemma shows that for the special class of matrices Λ arising in our case, there is a solution to the homogeneous linear system $\Lambda \boldsymbol{p} = \boldsymbol{0}$ that defines a probability distribution.

Lemma 3.2. *Let A be a $n \times (n+1)$ matrix defined as $A_{i,j} = a_{i-j}$, where $a_{-n}, \ldots, a_{n-1}$ is a sequence of numbers, with $a_0, \ldots, a_{n-1} \geq 0$, such that $a_{m-1} = -a_{-m}$, for all $m \geq -n$, and $a_{m-1} \geq a_m$ for all $m \geq 1$. Then, the system $A \boldsymbol{p} = \boldsymbol{0}$ has a symmetric solution with $p_j = p_{n-j}$, $\sum p_j = 1$, and $p_j \geq 0$. Moreover, there is a unique symmetric solution $\boldsymbol{p}$ that satisfies these conditions.*

Proof. By the hypothesis of the lemma, the matrix A has the following form:

$$A = \begin{pmatrix} a_0 & -a_0 & -a_1 & \ldots & -a_{n-1} \\ a_1 & a_0 & -a_0 & \ldots & -a_{n-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n-1} & \ldots & \ldots & \ldots & -a_0 \end{pmatrix}$$

8

We let $d_m = a_{m-1} - a_m \geq 0$, for any $m \geq 1$. Then, taking the difference of every pair of A's consecutive rows, we obtain the $(n-1) \times (n+1)$ matrix

$$A' = \begin{pmatrix} -d_1 & 2a_0 & -d_1 & \ldots & -d_{n-1} \\ -d_2 & -d_1 & 2a_0 & \ldots & -d_{n-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -d_{n-2} & -d_{n-3} & \ldots & -d_1 & -d_2 \\ -d_{n-1} & -d_{n-2} & \ldots & 2a_0 & -d_1 \end{pmatrix}$$

To establish the lemma, we first use Lemma 3.1 and show that (i) the nullspace of A' contains a unique symmetric probability vector $\boldsymbol{p}$, and then show that (ii) the particular vector $\boldsymbol{p}$ is also in the nullspace of A.

As for claim (i), we first show that each coordinate p_j of any vector $\boldsymbol{p}$ in the nullspace of A' can be expressed as a non-negative linear combination of the coordinates p_0 and p_n. Formally, we show that for any coordinate p_j of any solution $\boldsymbol{p}$ of $A'\boldsymbol{p} = \boldsymbol{0}$, there exist $\pi_j, \rho_j \geq 0$, such that $p_j = \pi_j p_0 + \rho_j p_n$. To this end, we consider the $(n-1) \times (n+1)$ matrix A'', which is obtained from A' by moving the first column of A' to the end. We observe that A'' satisfies the conditions of Lemma 3.1, since $\sum_{m=1}^{n-1} d_m = a_0 - a_{n-1} < a_0$, and thus $2a_0 - 2\sum_{m=1}^{n-1} d_m > 0$. Hence, by applying Gaussian elimination to A'', we get a $(n-1) \times (n+1)$ matrix G in a row-echelon form with $G_{i,i} > 0$, for all i, $G_{i,j} = 0$, for all $i > j$, and $G_{i,j} \leq 0$, for all $i < j$. Moreover, the nullspace of A' essentially consists of the solutions $\boldsymbol{x}$ to the homogenous linear system $G\boldsymbol{x} = \boldsymbol{0}$. More precisely, any solution $\boldsymbol{x}$ of $G\boldsymbol{x} = \boldsymbol{0}$ corresponds to a solution $\boldsymbol{p}$ of $A'\boldsymbol{p} = \boldsymbol{0}$, where $p_0 = x_n$, $p_1 = x_0$, $\ldots$, $p_n = x_{n-1}$, and vice versa.

Due to the special form of G, we can find all solutions $\boldsymbol{x}$ of $G\boldsymbol{x} = \boldsymbol{0}$ by assigning values to the free variables x_{n-1} and x_n and performing backwards substitution so that we uniquely determine the values of the variables $x_0, \ldots, x_{n-2}$. Furthermore, due to the special form of G, this procedure results in expressing each variable x_j as a non-negative linear combination of x_{n-1} and x_n. Specifically, we can calculate x_j, for all $j = n - 2, \ldots, 0$, from the equation $\sum_{i=0}^{n} G_{j,i} x_i = 0$. Solving for x_j, we get $x_j = -\sum_{i=j+1}^{n} G_{j,i} x_i / G_{j,j}$, since $G_{j,j} > 0$ and $G_{j,i} = 0$, for all $j > i$. Moreover, all coefficients $-G_{j,i}/G_{j,j}$ are non-negative because $G_{j,i} \leq 0$, for all $j < i$, and $G_{j,j} > 0$. By induction, if every $x_{j'}$, $j' > j$, is a non-negative linear combination of x_{n-1} and x_n, the same holds for x_j. Therefore, any coordinate x_j of any solution $\boldsymbol{x}$ to $G\boldsymbol{x} = \boldsymbol{0}$ can be expressed as a non-negative linear combination of the free variables x_{n-1} and x_n. Due to the aforementioned correspondence between the solutions $\boldsymbol{p}$ of $A'\boldsymbol{p} = \boldsymbol{0}$ and the solutions $\boldsymbol{x}$ of $G\boldsymbol{x} = \boldsymbol{0}$, we obtain that for any coordinate p_j of any solution $\boldsymbol{p}$ to $A'\boldsymbol{p} = \boldsymbol{0}$, there exist $\pi_j, \rho_j \geq 0$, such that $p_j = \pi_j p_0 + \rho_j p_n$.

Hence, the nullspace of A' is spanned by the vectors $\boldsymbol{p}^1$ and $\boldsymbol{p}^2$ determined by setting the free variables p_0 and p_n to $(1, 0)$ and to $(0, 1)$, respectively. By the discussion above, all the coordinates of $\boldsymbol{p}^1$ and $\boldsymbol{p}^2$ are non-negative. To conclude the proof of claim (i), we observe that due to the symmetry of the homogeneous linear system $A'\boldsymbol{p} = \boldsymbol{0}$, we have that $p_j^1 = p_{n-j}^2$, for all $j = 0, \ldots, n$. More specifically, by the definition of A', for all $i = 0, \ldots, n - 2$, the dot product of A''s i-th row and a vector $(a_0, a_1, \ldots, a_{n-1}, a_n)$ is identical to the dot product of A''s $(n-2-i)$-th row and the symmetric vector $(a_n, a_{n-1}, \ldots, a_1, a_0)$. Hence, $A'(a_0, \ldots, a_n) = \boldsymbol{0}$ if and only if $A'(a_n, \ldots, a_0) = \boldsymbol{0}$ for all vectors $(a_0, \ldots, a_n)$. Since $A'\boldsymbol{p}^1 = \boldsymbol{0}$, with $\boldsymbol{p}^1 = (\pi_0, \ldots, \pi_n)$, we have that $A'(\pi_n, \ldots, \pi_0) = \boldsymbol{0}$. Using that $\boldsymbol{p}^1$ and $\boldsymbol{p}^2$ are linearly independent and comprise a basis of the nullspace of A', we conclude that $\boldsymbol{p}^2 = (\pi_n, \ldots, \pi_0)$. Therefore, there is a unique symmetric vector in the nullspace of A' with L_1 norm equal to 1, namely the vector $\boldsymbol{p} = (\boldsymbol{p}^1 + \boldsymbol{p}^2)/|\boldsymbol{p}^1 + \boldsymbol{p}^2|_1$.

9

We proceed to show claim (ii), namely that the unique symmetric probability vector $\boldsymbol{p}$ in the nullspace of A' is also in the nullspace of A. To this end, we define the $n \times n$ matrix

$$M = \begin{pmatrix} 1 & 0 & 0 & \ldots & 1 \\ -1 & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \ldots & -1 & 1 & 0 \\ 0 & \ldots & 0 & -1 & 1 \end{pmatrix}$$

We observe that the determinant of M is equal to 2, and thus M is non-singular. Therefore, the linear system $A\boldsymbol{p} = \boldsymbol{0}$ is equivalent to the linear system $MA\boldsymbol{p} = \boldsymbol{0}$. So, we let A_1 and A_n be the first and the last row of A, and further observe that MA is a $n \times (n+1)$ matrix with its first row equal to $A_1 + A_n$ and its remaining rows in one-to-one correspondence to the rows of A'. Since $\boldsymbol{p}$ is the unique symmetric probability vector satisfying $A'\boldsymbol{p} = \boldsymbol{0}$, we only need to show that $(A_1 + A_n)\boldsymbol{p} = 0$, which follows immediately from the symmetry of $\boldsymbol{p}$. This completes the proof of claim (ii) and the proof of the lemma. $\square$

For every ℓ, the homogeneous linear system $\Lambda\boldsymbol{p} = \boldsymbol{0}$ satisfies the conditions of Lemma 3.2. Hence, there exists a unique symmetric probability distribution $\boldsymbol{p}$ such that the expected cost $\mathbb{E}[c(|X(\ell)-x|)]$ is the same for every location $x \in [0, \ell]$. Next, we think of this unique symmetric solution $\boldsymbol{p}$ as a function of ℓ, and establish a nice continuity property of it.

To this end, we fix an integer $m \geq 0$, and show that the random variable $X(\ell)$ converges in probability to the random variable $X(m)$, as $\ell \to m^+$. We observe that the linear system determining $\boldsymbol{p}$ is the same for all $\ell \in (m, m+1)$. So, we let p_i^m be the probability assigned to each integer point $i, 0 \leq i \leq m$. By symmetry, the probability assigned to each point $\ell - i, 0 \leq i \leq m$, is also p_i^m. The limit $\lim_{\ell \to m^+} X(\ell) = \bar{X}$ is a random variable distributed according to a probability distribution that assigns probability $p_i^m + p_{m-i}^m$ to each integer point $i, 0 \leq i \leq m$. Since the distribution is symmetric and achieves the same expected cost for all points $x \in [0, m]$, it is, by Lemma 3.2, the unique distribution with these properties. Therefore, we have that $X(m) = \bar{X}$. By the same argument, we can show that the random variable $X(\ell)$ converges in probability to the random variable $X(m+1)$, as $\ell \to (m+1)^-$.

By the continuity property above, the expected cost $C(\ell) = \mathbb{E}[c(X(\ell))]$ at each location $x \in [0, \ell]$ is a continuous function of ℓ. Moreover, the discussion above implies that for all $\ell \in [m, m+1)$, $C(\ell) = \sum_{i=0}^{m} p_i^m(c(i) + c(\ell - i))$. Using these properties, we now show that $C(\ell)$ is an increasing function of ℓ.

Lemma 3.3. *The expected cost $C(\ell)$ is an increasing function of the interval length ℓ.*

Proof. Since $C(\ell)$ is continuous, we only need to show that C is increasing in each interval $[m, m+1)$, where $m \geq 0$ is any integer. To this end, we let $\ell \in [m, m+1)$, and consider any $\ell' \in (\ell, m+1)$. Then, we have that:

$$C(\ell) = E[c(X(\ell))] = \sum_{i=0}^{m} p_i^m(c(i) + c(\ell - i)) < \sum_{i=0}^{m} p_i^m(c(i) + c(\ell' - i)) = C(\ell'),$$

where the inequality holds because $\ell' > \ell$ and the cost function c is increasing. $\square$

3.3 Step 3: Establishing Group Strategyproofness

We next prove that the random facility placement, in Step 3 of EQUAL COST, is group strategyproof. The correlation of the facility placement, in Step 3, ensures that if an agent j is located at y, his closest facility is always the one assigned to his closest interval. To justify this, let us consider any sample x of the random variable X. We recall that the facilities are placed at $\alpha_1 + x, \alpha_2 + \ell - x, \alpha_3 + x, \ldots$. Let us assume that $\alpha_i + \ell - x \leq y \leq \alpha_{i+1} + x$. Then, the distance of y to $\alpha_i + \ell - x$ is $y - (\alpha_i + \ell - x)$, while the distance of y to $\alpha_{i+1} + x$ is $\alpha_{i+1} + x - y$. Hence, agent j prefers the facility at interval i if and only if $y - (\alpha_i + \ell) < \alpha_{i+1} - y$, i.e., the right endpoint of interval i is closer to y than the left endpoint of interval $i + 1$.

To show that EQUAL COST is group strategyproof, we consider a coalition of agents S that deviate to improve their cost. Let the original interval length, with respect to the true agents' locations, be ℓ, and let the new interval length, after the deviation, be ℓ'. We now consider the two possible outcomes when the agents misreport their locations:

Case where $\ell' \geq \ell$. Let i be any agent. If i's true location is covered by some interval of the new covering, i incurs an expected cost of $C(\ell') \geq C(\ell)$. Otherwise, agent i incurs an expected cost no less than $C(\ell')$, which is greater than $C(\ell)$.

Case where $\ell' < \ell$. We consider the distance of any agent to the nearest midpoint of an interval. The locations of the truthful agents in $N \setminus S$ are covered by some interval of the new covering. Hence, their distance to the nearest midpoint of some interval is at most $\ell'/2$. On the other hand, if we consider the true locations of all agents and any feasible covering of them with k intervals, there is some agent whose distance to the midpoint of the interval covering him is at least $\ell/2$. Therefore, there is an agent i whose distance d to the nearest midpoint of some interval in the new covering (after the deviation) is at least $\ell/2$. Hence, agent i must be in the deviating coalition S, and his true location must not be covered by the intervals of the new covering. In this case, Lemma 3.4 below implies that the expected cost of agent i after the deviation, which is $\mathbb{E}[c(d - \ell'/2 + X(\ell'))]$, is at least as large as $\mathbb{E}[c(X(2d))] = C(2d) \geq C(\ell)$. This implies that EQUAL COST is group strategyproof.

Lemma 3.4. *For all a, a', b, with $0 \leq a < a' \leq b$, it holds that*

$$\mathbb{E}[c(b - a + X(2a))] \geq \mathbb{E}[c(b - a' + X(2a'))]$$

Moreover, the inequality is strict, if the function c is strictly concave.

Proof. Let $m \geq 0$ be any integer. We observe that, similarly to $C(\ell)$, $\mathbb{E}[c(b - a + X(2a))]$ is a continuous decreasing function of a. Therefore, we only need to show that the lemma holds for all $a, a' \in [\frac{m}{2}, \frac{m+1}{2})$, with $0 \leq a < a' \leq b$. For all such a, a', b, we have that:

$$\mathbb{E}[c(b - a + X(2a))] = \sum_{i=0}^{m} p_i^m (c(b - a + i) + c(b + a - i))$$

$$\geq \sum_{i=0}^{m} p_i^m (c(b - a' + i) + c(b + a' - i)) = \mathbb{E}[c(b - a' + X(2a'))]$$

where the inequality holds because $a < a'$ and c is concave. In fact, the inequality is strict if c is strictly concave. $\square$

11

3.4 Approximation Ratio

In this section, we analyze the approximation ratio of EQUAL COST.

Lemma 3.5. *For any concave cost function c, any locations profile $\boldsymbol{x}$, and any agent i, it holds that* $\mathrm{cost}(x_i, \mathrm{EC}(\boldsymbol{x}, c)) \leq \mathrm{MC}^*(\boldsymbol{x}, c)$.

Proof. We let ℓ be the minimum interval length in Step 1 of EQUAL COST, and let $m = \lfloor \ell \rfloor$. We recall that $\mathrm{MC}^*(\boldsymbol{x}, c) = c(\ell/2)$. Moreover, we have that:

$$C(\ell) = \sum_{i=0}^{m} p_i^m (c(i) + c(\ell - i)) \leq \sum_{i=0}^{m} 2 p_i^m c(\ell/2) = c(\ell/2)$$

where the inequality follows from the concavity of the cost function c. □

Lemma 3.6. *For every concave cost function c, EQUAL COST has an approximation ratio of at most 2 for the objective of* MAX COST.

Proof. Let $(\boldsymbol{x}, c)$ be any instance with a concave cost function c, and let ℓ be the minimum interval length in Step 1 of EQUAL COST. In $\mathrm{EC}(\boldsymbol{x}, c)$, every agent i has a facility at distance at most ℓ to x_i. On the other hand, $\mathrm{MC}^*(\boldsymbol{x}, c) = c(\ell/2)$. Therefore, the approximation ratio is at most:

$$\frac{c(\ell)}{c(\ell/2)} = \frac{c(\ell) + c(0)}{c(\ell/2)} \leq \frac{2c(\ell/2)}{c(\ell/2)} = 2,$$

where we use that $c(0) = 0$, by normalization, and the concavity of c. □

Lemma 3.7. *For every concave cost function c, EQUAL COST has an approximation ratio of at most n for the objective of* SOCIAL COST.

Proof. For every locations profile $\boldsymbol{x}$, $\mathrm{MC}^*(\boldsymbol{x}, c) \leq \mathrm{SC}^*(\boldsymbol{x}, c)$. Then,

$$\mathrm{SC}(\boldsymbol{x}, c) = \sum_{i \in N} \mathrm{cost}(x_i, \mathrm{EC}(\boldsymbol{x}, c)) \leq n\, \mathrm{MC}^*(\boldsymbol{x}, c) \leq n\, \mathrm{SC}^*(\boldsymbol{x}, c),$$

where the inequality follows from Lemma 3.5. □

4 Applications

In this section, we consider three typical examples of concave cost functions, and derive closed form solutions for the corresponding random variables $X(\ell)$.

Linear Functions. The literature mostly focuses on linear cost functions $c(d) = \lambda d$, where the agents' cost is proportional to their distance to the nearest facility. In this case, $X(\ell)$ has a nice closed form: it is either 0 with probability $1/2$ or ℓ with probability $1/2$. Then, the expected connection cost of any location $x \in [0, \ell]$ is:

$$c(x)/2 + c(\ell - x)/2 = \lambda x/2 + \lambda(\ell - x)/2 = 2\lambda \ell/2,$$

which does not depend on x.

Two-Piece Piecewise Linear Functions. For some $\lambda_1 > \lambda_2 > 0$, let the cost function c be:

$$c(d) = \begin{cases} \lambda_1 d & \text{for } d \le 1 \\ \lambda_2 d + (\lambda_1 - \lambda_2) & \text{for } d > 1 \end{cases}$$

To achieve the same expected cost at all locations, we find ℓ, let $m = \lfloor \ell \rfloor$, and compute the probability distribution of $X(\ell)$ by solving the following linear system:

$$\begin{pmatrix} \lambda_1 & -\lambda_1 & -\lambda_1 & -\lambda_2 & -\lambda_2 & -\lambda_2 & -\lambda_2 & \ldots & -\lambda_2 & -\lambda_2 & -\lambda_2 \\ \lambda_1 & \lambda_1 & -\lambda_1 & -\lambda_1 & -\lambda_2 & -\lambda_2 & -\lambda_2 & -\lambda_2 & \ldots & -\lambda_2 & -\lambda_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \lambda_2 & \lambda_2 & \lambda_2 & \ldots & \ldots & \ldots & \lambda_2 & \lambda_2 & \lambda_1 & \lambda_1 & -\lambda_1 \end{pmatrix} \begin{pmatrix} p_0^m \\ p_m^m \\ p_1^m \\ p_{m-1}^m \\ \vdots \\ p_m^m \\ p_0^m \end{pmatrix} = 0$$

Taking the difference between every two consecutive rows, as in Lemma 3.2, we find that:

$$p_i^m = \frac{\lambda_1 - \lambda_2}{2\lambda_1}(p_{i-1}^m + p_{i+1}^m) \quad \text{for all integers } i \ 0 \le i \le m \, ,$$

where we define $p_i^m = 0$, for all integers $i \notin [0, m]$. Then, the solution of the recurrence is:

$$p_i^m = \frac{\rho_1^{m+1-i} + \rho_2^{m+1-i}}{2 \sum_{j=1}^{m+1} \left(\rho_1^j + \rho_2^j \right)}$$

$$\text{where} \quad \rho_1 = \frac{\lambda_1 + \sqrt{\lambda_1^2 - (\lambda_1 - \lambda_2)^2}}{\lambda_1 - \lambda_2} \quad \text{and} \quad \rho_2 = \frac{\lambda_1 - \sqrt{\lambda_1^2 - (\lambda_1 - \lambda_2)^2}}{\lambda_1 - \lambda_2}$$

Exponential Functions. A concave cost function that results in a continuous probability distribution $X(\ell)$ is the exponential function $c(d) = 1 - e^{-\lambda d}$. Then, $X(\ell)$ is 0, with probability $\frac{1}{\ell\lambda+2}$, ℓ, with probability $\frac{1}{\ell\lambda+2}$, and uniform in $(0, \ell)$, with probability $\frac{\ell\lambda}{\ell\lambda+2}$. Then, the expected connection cost of any location $x \in [0, \ell]$ is equal to $\frac{\ell\lambda}{\ell\lambda+2}$, which does not depend on x.

5 Extensions and Limitations

5.1 EQUAL COST **in Bounded Intervals**

Our results about the properties of EQUAL COST apply to the real line $(-\infty, \infty)$ and to the half-line $[0, \infty)$. If the metric space is a bounded interval $[0, L]$, it could be that in the construction of the covering, in Step 1, the last interval does not fit entirely in $[0, L]$. The following lemma shows that even in this case, we can adjust the covering with disjoint intervals of the same length, computed in Step 1, so that all intervals fit in $[0, L]$.

Lemma 5.1. *Given a locations profile x in $[0, L]$, there is an optimal covering of x with k disjoint intervals of the same (minimum) length that all lie entirely in $[0, L]$.*

Proof. We consider a covering of $\boldsymbol{x}$ with k disjoint intervals of the same minimum length ℓ, computed as in Section 3.1. As in Step 1 of EQUAL COST, we number the intervals from left to right, and let the i-th interval be $[\alpha_i, \alpha_i + \ell]$. Since all the locations of $\boldsymbol{x}$ lie in $[0, L]$, we obtain that $\ell \leq L/k$. Moreover, by construction, we have that $\alpha_i \geq 0$, for all $1 \leq i \leq k$. However, it could be $\alpha_i + \ell > L$ for some interval i. In this case, we construct a new covering using the intervals $[\alpha_i', \alpha_i' + \ell]$, $i = 1, \ldots, k$, where $\alpha_i' = \min\{\alpha_i, L - (k + 1 - i)\ell\}$. To show that this is indeed an admissible covering of $\boldsymbol{x}$, we observe that:

(i) All intervals lie entirely in $[0, L]$: For every i, $\alpha_i' \geq 0$, since $\alpha_i \geq 0$, and $L - (k + 1 - i)\ell \geq 0$, because $\ell \leq L/k$. Furthermore, $\alpha_i' + \ell \leq L - (k + 1 - i)\ell + \ell \leq L - (k - i)\ell \leq L$.

(ii) All intervals are disjoint: For any two consecutive intervals i and $i + 1$, we have that:

$$\begin{aligned}
\alpha_{i+1}' - \alpha_i' &= \min\{\alpha_{i+1}, L - (k + 1 - i - 1)\ell\} - \min\{\alpha_i, L - (k + 1 - i)\ell\} \\
&\geq \min\{\alpha_i + \ell, L - (k + 1 - i - 1)\ell\} - \min\{\alpha_i, L - (k + 1 - i)\ell\} \\
&= \ell + \min\{\alpha_i, L - (k + 1 - i)\ell\} - \min\{\alpha_i, L - (k + 1 - i)\ell\} \\
&= \ell
\end{aligned}$$

(iii) The intervals cover all locations of $\boldsymbol{x}$: Let us consider a location $x \in [\alpha_i, \alpha_i + \ell]$. If $\alpha_i' = \alpha_i$, x is covered since the interval does not change. Otherwise, $\alpha_i' = L - (k + 1 - i)\ell$. Thus, the interval $[\alpha_i', L]$ has a length of $L - \alpha_i' = (k + 1 - i)\ell$, and consists of $k + 1 - i$ disjoint intervals of length ℓ. Therefore, the intervals $[\alpha_j', \alpha_j' + \ell]$, for $j \geq i$, entirely cover the interval $[\alpha_i', L] \supset [\alpha_i, L]$, and thus, they also cover the location x. □

Lemma 5.1 implies that if the agents lie on a circle, we can also cover their locations with disjoint intervals of the same minimum length ℓ. Then, we can apply Steps 2 and 3 to the resulting intervals on the circle. But rather surprisingly, EQUAL COST is guaranteed to be strategyproof for k-Facility Location on the circle only if k is even. Otherwise, some agents in the first interval may prefer the facility placed in the last interval, which violates the property that each agent always prefers the facility in his own interval.

5.2   Convex Cost Functions

The approach of EQUAL COST does not apply to strictly convex functions c, because it is no longer possible to equalize the expected cost of all agents. To see this, let us consider the interval $[0, \ell]$, and the expected cost of two agents, one located at 0 and the other at ℓ. Since $\mathbb{E}[c(X)] + \mathbb{E}[c(\ell - X)] > \mathbb{E}[2c(\ell/2)] = 2c(\ell/2)$, by the strict convexity of c, at least one of them incurs an expected cost greater than $c(\ell/2)$. However, a third agent located at $\ell/2$ incurs an expected cost no greater than $c(\ell/2)$, since his distance to the facility is at most $\ell/2$. Moreover, we can show that:

Lemma 5.2. *There is no randomized strategyproof mechanism that achieves a bounded approximation ratio for the class of all convex functions.*

Proof. We recall that the property of a bounded approximation ratio is objective-independent. So, we next focus on the objective of MAX COST. For the proof, we consider the convex cost function $c(d) = e^d$ and instances with 2 agents and a single facility. For sake of contradiction, we assume that there exists a randomized strategyproof mechanism that achieves an approximation ratio of r for such

instances. Next, we let X denote the random variable that determines where the mechanism places the facility.

We first consider an instance $\boldsymbol{x} = (x_1, x_2)$, with $x_2 > x_1$, If the facility is placed at location $t \leq (x_1 + x_2)/2$, agent 2 incurs the maximum cost equal to $e^{x_2 - t}$. If the facility is placed at $t > (x_1 + x_2)/2$, agent 1 incurs the maximum cost equal to $e^{t - x_1}$. In both cases, the maximum cost is equal to $e^{(x_2 - x_1)/2 + |t - (x_1 + x_2)/2|}$, and the expectation of the maximum cost is $\mathbb{E}[e^{(x_2 - x_1)/2 + |X - (x_1 + x_2)/2|}] \leq r e^{(x_2 - x_1)/2}$, which implies that $\mathbb{E}[e^{|X - (x_1 + x_2)/2|}] \leq r$.

Let us now consider the probabilities $p_l = \Pr[X \leq \frac{x_1 + x_2}{2}]$ and $p_r = \Pr[X \geq \frac{x_1 + x_2}{2}]$. Since $p_l + p_r \geq 1$, one of them is at least $1/2$. Wlog., let us assume that $p_l \geq 1/2$, which implies that agent 2 incurs an expected cost of at least $\frac{1}{2} e^{(x_2 - x_1)/2}$.

Next, we consider an instance $\boldsymbol{x}' = (x_1', x_2')$, with $x_1' = x_1$ and $x_2' = 2x_2 - x_1$. By the choice of $\boldsymbol{x}'$, $\mathbb{E}[e^{|X - (x_1' + x_2')/2|}] = \mathbb{E}[e^{|X - (x_1 + 2x_2 - x_1)/2|}] = \mathbb{E}[e^{|X - x_2|}]$. Working as before, we obtain that $\mathbb{E}[e^{|X - (x_1' + x_2')/2|}] = \mathbb{E}[e^{|X - x_2|}] \leq r$, due to the approximation ratio of the mechanism. Moreover, $\mathbb{E}[e^{|X - x_2|}]$ is the expected cost of an agent located at x_2, and due to strategyproofness, is no less than the expected cost of agent 2 in instance $\boldsymbol{x}$. Otherwise agent 2 would have an incentive to report x_2', instead of x_2. Therefore, $\mathbb{E}[e^{|X - x_2|}] \geq \frac{1}{2} e^{(x_2 - x_1)/2}$. Combining the upper and the lower bound on $\mathbb{E}[e^{|X - x_2|}]$, we obtain that $e^{(x_2 - x_1)/2} \leq 2r$. This leads to a contradiction if we consider an instance $\boldsymbol{x}$ with $x_2 - x_1 > 2\ln(2r)$. $\qquad\square$

5.3 Other Cost Functions

EQUAL COST can also apply to some other (non-convex) cost functions, for which the expected cost of all agents can be equalized. A notable such example is a cost function $c_r(d)$ which is 0, if $d < r$, and 1 otherwise. Thus, c_r correspond to agents that only care about getting a facility within a radius r from their location. In this case, one could apply EQUAL COST as follows: First, we find a covering of the agent locations with intervals of length ℓ, as in Step 1. Then if $\ell \leq 2r$, we place a facility at the midpoint of each interval. Otherwise, we do not place any facilities (and let each agent incur a cost of 1). This clearly satisfies the equal cost property since the cost incurred by all agents is either 0 or 1. The mechanism is optimal for the objective of MAX COST because every agent incurs a cost of 0, if the optimal solution satisfies all agents, and a cost of 1, otherwise. On the other hand, the mechanism is n approximate for the objective of SOCIAL COST, since in case where the optimal solution satisfies all but one agents, resulting in a social cost of 1, the mechanism does not place any facilities, and incurs a social cost of n.

6 The PICK THE LOSER Mechanism

EQUAL COST performs well for the objective of MAX COST, but may perform poorly for the objective of SOCIAL COST. An extreme case is when we have k facilities and only $n = k + 1$ agents. Then, there are many facilities, and one could easily satisfy all but one agents. Nevertheless, EQUAL COST causes all agents to incur a high cost (equal to the min-max cost for linear cost functions).

In certain cases, this might not be acceptable, and one needs to find a more efficient mechanism. In this section, we present a mechanism that, for instances with only $n = k + 1$ agents, selects the loser, i.e., the agent not allocated a facility at his location, in a group strategyproof way. We also show that this mechanism is quite efficient for the SOCIAL COST objective, which for such instances, is equal to the cost of the loser.

Given an instance $(\boldsymbol{x}, c)$ of k-Facility Location on the line with only $n = k + 1$ agents, the PICK-THE-LOSER mechanism, of PtL in short, works as follows:

Step 1 It numbers the agents according to their reported locations such that $x_i < x_{i+1}$, and lets E and O be the sets of even and odd numbered agents, respectively. For every odd-numbered agent $i \in O$, PtL places a facility at x_i.

Step 2 For each even numbered agent i, PtL samples a number s_i uniformly in $(0, 1)$, and computes i's current cost $\kappa_i = \min_{j \neq i} c(|x_j - x_i|)$ and i's scaled cost $\hat{\kappa}_i = \kappa_i / s_i$.

Step 3 PtL finds the agent with the smallest scaled cost, and declares him the *loser*. Then, PtL places facilities at the locations of all other agents.

In the following, we first show that PICK THE LOSER is strategyproof (Lemma 6.1). Then, in Section 6.2, we use strategyproofness, and deal with the case where a coalition of agents may deviate, thus establishing that the mechanism is strongly group strategyproof. Finally, in Section 6.3, we prove the mechanism's approximation guarantee. Thus, we obtain:

Theorem 6.1. *For the class of all concave cost functions,* PICK THE LOSER *is strongly group strategyproof and achieves an approximation ratio of* 2 *for the* SOCIAL COST *objective.*

For the proof, we assume wlog. that the agent locations are all distinct. Otherwise, we allocate a facility to all distinct locations, thus being trivially both optimal and group strategyproof. We let $q_i(\boldsymbol{x})$ denote the probability that agent i is designated as a loser. We have that $q_i(\boldsymbol{x}) = 0$ for all odd numbered agents in $\boldsymbol{x}$. For an even numbered agent i, we can compute this probability by the following thought experiment: With all the samples $s_j \in (0, 1)$ fixed, agent i is selected if for all $j \in E$, $\hat{\kappa}_j \geq \hat{\kappa}_i$, or equivalently if $s_j \leq s_i \kappa_j / \kappa_i$. This happens with probability $\prod_{j \in E \setminus \{i\}} \min\{1, s_i \kappa_j / \kappa_i\}$. Setting $t = s_i / \kappa_i$ and taking the expectation over all different values of t, we have that

$$q_i(\boldsymbol{x}) = \kappa_i \int_0^{1/\kappa_i} \prod_{j \in E \setminus \{i\}} \min\{1, \kappa_j t\} dt$$

6.1 Strategyproofness

The following lemma implies that PICK THE LOSER is strategyproof for the class of all concave cost functions. Next, in Section 6.2, we use this property, to establish that PICK THE LOSER is strongly group strategyproof for the class of all concave cost functions.

Lemma 6.1. *Let $(\boldsymbol{x}, c)$ be any instance with a concave cost function c and only $n = k+1$ agents occupying n distinct locations. Then, for every agent i and every location $x_i' \neq x_i$, $\mathrm{cost}(x_i, \mathrm{PtL}(\boldsymbol{x}, c)) < \mathrm{cost}(x_i, \mathrm{PtL}((\boldsymbol{x}_{-i}, x_i'), c))$.*

Proof. For convenience, we let $\boldsymbol{x}' = (\boldsymbol{x}_{-i}, x_i')$. We also recall that by normalizing c, we assume that $c(0) = 0$. If agent i is an odd numbered agent, he strictly prefers $\boldsymbol{x}$ over $\boldsymbol{x}'$, because in $\boldsymbol{x}$, there is a facility at x_i and agent i incurs 0 cost, while in x_i', there is no facility at x_i, and thus agent i incurs a positive cost.

If i is an even numbered agent, we let $\delta = \min_{j \neq i}\{|x_i - x_j|\}$ and $\delta' = \min_{j \neq i}\{|x_i' - x_j|\}$ denote the minimum distance of the reported location of i to the location of another agent. In the instance $\boldsymbol{x}$, if agent i is not allocated a facility at x_i, he incurs a cost of $c(\delta)$. Otherwise, agent i incurs 0 cost.

Since $\delta > 0$ and the cost function is increasing in c, we have that $c(\delta) > 0$. We next consider three different cases, and show that in each case, agent i prefers $\boldsymbol{x}$ to $\boldsymbol{x}'$.

Case where $x_i' \notin (x_i - \delta, x_i + \delta)$. Then, in $\boldsymbol{x}'$, agent i incurs an expected cost of at least $c(\delta)$, while in $\boldsymbol{x}$, he incurs an expected cost less than $c(\delta)$, since he is allocated a facility at x_i with positive probability.

Case where $x_i' \in (x_i - \delta, x_i + \delta)$ and $\delta' \leq \delta$. In this case, the probability $q_i(\boldsymbol{x}')$ that agent i is not allocated a facility at x_i', in instance $\boldsymbol{x}'$, is greater than or equal to $q_i(\boldsymbol{x})$. This holds because i's cost in $\boldsymbol{x}'$, which is $\kappa_i' = c(\delta')$, is less than or equal to i's cost in $\boldsymbol{x}$, which is $\kappa_i = c(\delta)$. Therefore, for any sampled number s_i, agent i has a smaller scaled cost $\hat{\kappa}_i'$ in instance $\boldsymbol{x}'$ than his corresponding scaled cost $\hat{\kappa}_i$ in instance $\boldsymbol{x}$, which in turn, implies a greater probability that i is designated as the loser. Moreover, if in instance $\boldsymbol{x}'$, agent i is allocated a facility at x_i', he incurs a positive cost, since x_i' is different from his true location x_i. Thus, putting everything together, we obtain that agent i strictly prefers $\boldsymbol{x}$ to $\boldsymbol{x}'$:

$$\left(1 - q_i(\boldsymbol{x}')\right) c(|x_i' - x_i|) + q_i(\boldsymbol{x}')c(\delta) > q_i(\boldsymbol{x}')c(\delta) \geq q_i(\boldsymbol{x})c(\delta)$$

Case where $x_i' \in (x_i - \delta, x_i + \delta)$ and $\delta' > \delta$. The probability $q_i(\boldsymbol{x}')$ is now greater than the probability $q_i(\boldsymbol{x})$. However, if agent i is allocated a facility at x_i', in instance $\boldsymbol{x}'$, he incurs an additional cost of $c(|x_i' - x_i|) \geq c(\delta' - \delta)$, due to the distance of x_i' to i's true location x_i. Thus, we obtain the following lower bound on the expected cost of agent i in instance $\boldsymbol{x}'$:

$$\left(1 - q_i(\boldsymbol{x}')\right) c(|x_i' - x_i|) + q_i(\boldsymbol{x}')c(\delta) \geq \left(1 - q_i(\boldsymbol{x}')\right) \left(c(\delta') - c(\delta)\right) + q_i(\boldsymbol{x}')c(\delta),$$

where the inequality follows from $c(\delta') \leq c(\delta) + c(\delta' - \delta)$, which in turn, follows from the concavity of c. Hence, to conclude that agent i strictly prefers $\boldsymbol{x}$ to $\boldsymbol{x}'$, we need to show that:

$$\left(1 - q_i(\boldsymbol{x}')\right) \left(c(\delta') - c(\delta)\right) + q_i(\boldsymbol{x}')c(\delta) > q_i(\boldsymbol{x})c(\delta) \tag{1}$$

To this end, for each even numbered agent j, we let κ_j and κ_j' denote the cost of j computed by the mechanism for the instances $\boldsymbol{x}$ and $\boldsymbol{x}'$, respectively. By the definition of the mechanism, we have that $\kappa_j = \kappa_j'$, for any agent $j \in E \setminus \{i\}$, and that $\kappa_i = c(\delta)$ and $\kappa_i' = c(\delta')$. Hence, the probability $q_i(\boldsymbol{x}')$ can be calculated as follows:

$$q_i(\boldsymbol{x}') = \kappa_i' \int_0^{1/\kappa_i'} \prod_{j \in E \setminus \{i\}} \min\{1, \kappa_j t\} dt \tag{2}$$

To prove (1), we show that i' expected cost is increasing with κ_i'. To prove this, we show that the partial derivative of i's cost with respect to κ_i' is positive. Formally, we show that:

$$\frac{\partial}{\partial \kappa_i'} \left[\left(1 - q_i(\boldsymbol{x}')\right) \left(\kappa_i' - \kappa_i\right) + q_i(\boldsymbol{x}')\kappa_i\right] > 0 \tag{3}$$

We first substitute $q_i(\boldsymbol{x}')$, with the use of (2), and the left-hand-side of (3) becomes:

$$\frac{\partial}{\partial \kappa_i'} \left[\left(1 - \kappa_i' \int_0^{1/\kappa_i'} \prod_{j \in E \setminus \{i\}} \min\{1, \kappa_j t\} dt\right) \left(\kappa_i' - \kappa_i\right) + \kappa_i \left(\kappa_i' \int_0^{1/\kappa_i'} \prod_{j \in E \setminus \{i\}} \min\{1, \kappa_j t\} dt\right)\right]$$

17

Next, we calculate the partial derivative with respect to κ_i', and the quantity above becomes:

$$1 - \prod_{j \in E \setminus \{i\}} \min\left\{1, \frac{\kappa_j}{\kappa_i'}\right\} + \frac{2(\kappa_i' - \kappa_i)}{\kappa_i'} \left(\prod_{j \in E \setminus \{i\}} \min\left\{1, \frac{\kappa_j}{\kappa_i'}\right\} - \kappa_i' \int_0^{1/\kappa_i'} \prod_{j \in E \setminus \{i\}} \min\{1, \kappa_j t\} dt \right)$$

Using that $\prod_{j \in E \setminus \{i\}} \min\{1, \kappa_j t\} \leq \prod_{j \in E \setminus \{i\}} \min\{1, \kappa_j/\kappa_i'\}$, which holds for all $t \in [0, 1/\kappa_i']$, and with strict inequality for $t < 1/\kappa_i'$, we obtain that the quantity above is greater than:

$$1 - \prod_{j \in E \setminus \{i\}} \min\left\{1, \frac{\kappa_j}{\kappa_i'}\right\} + \frac{2(\kappa_i' - \kappa_i)}{\kappa_i'} \left(\prod_{j \in E \setminus \{i\}} \min\left\{1, \frac{\kappa_j}{\kappa_i'}\right\} - \kappa_i' \prod_{j \in E \setminus \{i\}} \min\left\{1, \frac{\kappa_j}{\kappa_i'}\right\} \int_0^{1/\kappa_i'} 1 dt \right)$$

Simplifying the quantity above and returning back to (3), we conclude that:

$$\frac{\partial}{\partial \kappa_i'} \left[\left(1 - q_i(\boldsymbol{x}')\right) \left(\kappa_i' - \kappa_i\right) + q_i(\boldsymbol{x}')\kappa_i \right] > 1 - \prod_{j \in E \setminus \{i\}} \min\{1, \kappa_j/\kappa_i'\} \geq 0$$

Therefore, the expected cost of agent i is increasing with κ_i'. Hence, we obtain that

$$\left(1 - q_i(\boldsymbol{x}')\right) \left(\kappa_i' - \kappa_i\right) + q_i(\boldsymbol{x}')\kappa_i > q_i(\boldsymbol{x})\kappa_i \,,$$

which is identical to (1). This proves that in the third case, agent i strictly prefers $\boldsymbol{x}$ to $\boldsymbol{x}'$, and concludes the proof of the lemma. $\qquad \square$

6.2 Strong Group Strategyproofness

Proving that PICK THE LOSER is strong group strategyproof requires some additional arguments and case analysis, where we use that the mechanism is strategyproof (Lemma 6.1).

Throughout this section, we consider an instance $\boldsymbol{x}$ with n distinct locations, where the agents are numbered as they appear on the line, from left to right. Hence, we have that $x_i < x_{i+1}$, for all $i = 1, \ldots, n-1$. We prove that there is no coalition of agents that can benefit by misreporting their location. For sake of contradiction, let us assume that such a coalition exists. In particular, we let S be such coalition of minimum size, and let $\boldsymbol{x}' = (\boldsymbol{x}'_S, \boldsymbol{x}_{-S})$ be the new instance, where the agents in S misreport their location. By the definition of strong group strategyproofness, for all $i \in S$, $\mathrm{cost}(x_i, \mathrm{PtL}(\boldsymbol{x}', c)) \leq \mathrm{cost}(x_i, \mathrm{PtL}(\boldsymbol{x}, c))$, and the inequality is strict for at least one agent in S.

We observe that for every odd numbered agent i, $x_i \in \boldsymbol{x}'$. Otherwise, agent i would incur a positive cost in $\boldsymbol{x}'$, and would prefer $\boldsymbol{x}$ to $\boldsymbol{x}'$. We recall that PICK THE LOSER is anonymous, i.e., does not take the agent identities into account. Therefore, if $x_i' \neq x_i$, since some other agent j has reported the location x_i in $\boldsymbol{x}'$, we can change the identities of agents i and j without changing how the mechanism works. Hence, in all cases, we can assume wlog. that $x_i' = x_i$, which implies that the deviating coalition S doesn't contain any odd numbered agents.

Furthermore, we observe that for every even numbered agent i, there is a location in $\boldsymbol{x}'$ lying in the interval (x_{i-1}, x_{i+1}), where x_{n+1} is defined to be ∞. Otherwise agent i would incur an expected cost of $\mathrm{cost}(x_i, \mathrm{PtL}(\boldsymbol{x}', c)) \geq \min\{c(x_i - x_{i-1}), c(x_{i+1} - x_i)\}$, which is greater than his expected cost for instance $\boldsymbol{x}$, where x_i is allocated a facility with positive probability. Again, since PICK THE

LOSER is anonymous, we can assume wlog. that $x_i' \in (x_{i-1}, x_{i+1})$, which implies that the relative order of the agents in $\boldsymbol{x}'$ is the same as in $\boldsymbol{x}$.

Let us now consider an agent $i \in S$, and let κ_i and κ_i' denote the cost of i computed by the mechanism for the instances $\boldsymbol{x}$ and $\boldsymbol{x}'$, respectively. Next, we exclude the possibility that $\kappa_i' > \kappa_i$. Specifically, we show that if $\kappa_i' > \kappa_i$, the instance $\boldsymbol{x}'' = (\boldsymbol{x}_{-i}', x_i)$ is strictly preferable to $\boldsymbol{x}'$ for all agents in S. That holds because, in $\boldsymbol{x}''$, agent i has cost $\kappa_i < \kappa_i'$, and therefore, the probability that he is designated as the loser in $\boldsymbol{x}''$ is greater than the corresponding probability in $\boldsymbol{x}'$. Hence, for every agent $j \in E \setminus \{i\}$, the probability that agent j is designated as the loser in $\boldsymbol{x}''$ is less than the corresponding probability in $\boldsymbol{x}'$, which implies that agent j strictly prefers the instance $\boldsymbol{x}''$ to the instance $\boldsymbol{x}'$. Also by Lemma 6.1, PICK THE LOSER is strategyproof, and thus, agent i strictly prefers the instance $\boldsymbol{x}''$ to the instance $\boldsymbol{x}'$. However, since the number of agents misreporting their locations in $\boldsymbol{x}''$ is one less than the corresponding number in $\boldsymbol{x}'$, this contradicts the hypothesis that S is the smallest coalition of agents that can benefit from misreporting their location.

So, let us now assume that $\kappa_i' < \kappa_i$, for all agents $i \in S$, and let $\rho = \min\{\kappa_i/\kappa_i'\} > 1$. We consider an instance $\boldsymbol{x}''$ where the cost κ_i'' computed by the mechanism for all agents $i \in S$ is equal to $\rho \kappa_i'$. Such an instance $\boldsymbol{x}''$ can be obtained if we let all agents $i \in S$ report locations closer to their original location. We next show that for every agent $i \in S$, the probability $q_i(\boldsymbol{x}'')$ that i is designated as the loser in $\boldsymbol{x}''$ is less than the probability $q_i(\boldsymbol{x}')$ that i is designated as the loser in $\boldsymbol{x}'$. To this end, we consider below the probability that i has the smallest scaled cost $\hat{\kappa}_i''$, in $\boldsymbol{x}''$, and $\hat{\kappa}_i'$, in $\boldsymbol{x}'$, among all even numbered agents $j \notin S$, given that i has the smallest scaled cost among all agents in S. So, for clarity, the notation $j \notin S$ refers, in the calculation below, to all even numbered agents that do not belong to the coalition S. We have that:

$$
\begin{aligned}
q_i(\boldsymbol{x}'') &= \Pr[\hat{\kappa}_i'' < \hat{\kappa}_j'', \forall j \notin S \,|\, \hat{\kappa}_i'' < \hat{\kappa}_j'', \forall j \in S \setminus \{i\}] \cdot \Pr[\hat{\kappa}_i'' < \hat{\kappa}_j'', \forall j \in S \setminus \{i\}] \\
&= \Pr[\rho \hat{\kappa}_i' < \hat{\kappa}_j', \forall j \notin S \,|\, \rho \hat{\kappa}_i' < \rho \hat{\kappa}_j', \forall j \in S \setminus \{i\}] \cdot \Pr[\rho \hat{\kappa}_i' < \rho \hat{\kappa}_j', \forall j \in S \setminus \{i\}] \\
&= \Pr[\hat{\kappa}_i' < \hat{\kappa}_j'/\rho, \forall j \notin S \,|\, \hat{\kappa}_i' < \hat{\kappa}_j', \forall j \in S \setminus \{i\}] \cdot \Pr[\hat{\kappa}_i' < \hat{\kappa}_j', \forall j \in S \setminus \{i\}] \\
&< \Pr[\hat{\kappa}_i' < \hat{\kappa}_j', \forall j \notin S \,|\, \hat{\kappa}_i' < \hat{\kappa}_j', \forall j \in S \setminus \{i\}] \cdot \Pr[\hat{\kappa}_i' < \hat{\kappa}_j', \forall j \in S \setminus \{i\}] \\
&= q_i(\boldsymbol{x}')
\end{aligned}
$$

Since for every agent $i \in S$, (i) the probability that i is designated as the loser is smaller in $\boldsymbol{x}''$ than in $\boldsymbol{x}'$, i.e., $q_i(\boldsymbol{x}'') < q_i(\boldsymbol{x}')$, (ii) the reported location of i in $\boldsymbol{x}''$ is closer to his true location x_i than his reported location in $\boldsymbol{x}'$, i.e., $|x_i'' - x_i| < |x_i' - x_i|$, and (iii) there are no odd numbered agents in S, all agents $i \in S$ strictly prefer $\boldsymbol{x}''$ to $\boldsymbol{x}'$.

Therefore, we can assume that in the instance $\boldsymbol{x}'$, there is an agent $i \in S$ with $\kappa_i' = \kappa_i$. We now consider the instance $\boldsymbol{x}'' = (\boldsymbol{x}_{-i}', x_i)$, where the agent i is removed from the deviating coalition S. We note that for every agent $j \in S$, the probability that agent j is designated as the loser in $\boldsymbol{x}''$ is the same as the corresponding probability in $\boldsymbol{x}'$, i.e., $q_j(\boldsymbol{x}'') = q_j(\boldsymbol{x}')$. Therefore, the expected cost of every agent $j \in S \setminus \{i\}$ in $\boldsymbol{x}''$ is the same as his expected cost in $\boldsymbol{x}'$. Moreover, by Lemma 6.1, PICK THE LOSER is strategyproof, and thus, the expected cost of agent i in $\boldsymbol{x}''$ is less than his expected cost in $\boldsymbol{x}'$. Therefore, if the agents in the coalition S can benefit by misreporting their locations, the same holds for the $S \setminus \{i\}$. However, this contradicts the hypothesis that S is the smallest coalition of agents that can benefit from misreporting their location. Hence, we have shown that such a coalition S does not exist, and thus, the mechanism PICK THE LOSER is strongly group strategyproof.

6.3 Approximation Ratio

Lemma 6.2. *For all concave cost functions,* PICK THE LOSER *achieves an approximation ratio of at most* 2 *for the objective of* SOCIAL COST*, and an approximation ratio of at most* 4 *for the objective of* MAX COST*.*

Proof. Let $(\boldsymbol{x}, c)$ be any instance with concave c, and let q be an agent with $\kappa_q = \min_i\{\kappa_i\}$. Then, $\text{SC}^*(\boldsymbol{x}, c) = \kappa_q$, while the SOCIAL COST of the mechanism is equal to:

$$\sum_i \kappa_i \int_0^{1/\kappa_i} \kappa_i \prod_{j \in E \setminus \{i\}} \min\{1, \kappa_j t\} dt \leq \kappa_q + \sum_{i \neq q} \kappa_i \int_0^{1/\kappa_i} \kappa_i \prod_{j \in E \setminus \{i\}} \min\{1, \kappa_j t\} dt$$

$$= \kappa_q + \sum_{i \neq q} \kappa_i \int_0^{1/\kappa_i} \kappa_i \kappa_q t \prod_{j \in E \setminus \{i,q\}} \min\{1, \kappa_j t\} dt$$

$$\leq \kappa_q + \sum_{i \neq q} \kappa_i \int_0^{1/\kappa_i} \kappa_q \prod_{j \in E \setminus \{i,q\}} \min\{1, \kappa_j t\} dt$$

$$\leq \kappa_q + \kappa_q \sum_{i \neq q} \int_0^{1/\kappa_i} \kappa_i \prod_{j \in E \setminus \{i,q\}} \min\{1, \kappa_j t\} dt$$

$$= \kappa_q + \kappa_q \sum_{i \neq q} \Pr[\hat{\kappa}_i < \hat{\kappa}_j, \forall j \notin \{i, q\}]$$

$$= \kappa_q + \kappa_q \cdot 1 = 2\kappa_q$$

Moreover, since c is concave, $\text{MC}^*(\boldsymbol{x}, c) \geq \kappa_q/2 = 2\kappa_q/4 \geq \text{MC}(\text{PtL}(\boldsymbol{x}, c))/4$. $\qquad\square$

7 Open Problems

There are a few interesting open problems arising from our work. First, since EQUAL COST crucially depends on the linear structure of the instances, it would be interesting to have a mechanism that can be applied to more general metric spaces, and retains the nice properties of EQUAL COST. Another intriguing open problem has to do with the approximability of SOCIAL COST by randomized strategyproof mechanisms. Despite the considerable interest in the problem, we do not know whether there exists a randomized mechanism for k-Facility Location that achieves an approximation ratio of $o(n)$ for all $k \geq 3$. Another, more general, direction for further research may concern the role of the cost function c, which we assume here to be the same for all players. It would be interesting to investigate the approximability of k-Facility Location on the line if each agent i may have a different concave cost function $c_i(d)$. A good starting point in this direction may be a simple setting where each agent i is associated with a tuple (x_i, r_i), with possibly both x_i and r_i being private information, and there is some fixed small cost incurred by agent i, if there is a facility within a distance of r_i to x_i, and some fixed large cost incurred by agent i, otherwise.

References

1. N. Alon, M. Feldman, A.D. Procaccia, and M. Tennenholtz. Strategyproof approximation of the minimax on networks. *Mathematics of Operations Research*, 35(3):513-526, 2010.

2. D. Arthur and S. Vassilvitskii. *k*-means++: The advantages of careful seeding. In *Proc. of the 18th ACM-SIAM Symposium on Discrete Algorithms (SODA '07)*, pp. 1027-1035, 2007.

3. S. Barberà. An introduction to strategyproof social choice functions. *Social Choice and Welfare*, 18:619-653, 2001.

4. E. Dokow, M. Feldman, R. Meir, and I. Nehama. Mechanism design on discrete lines and cycles. In *Proc. of the 13th ACM Conference on Electronic Commerce (EC '12)*, pp. 423-440, 2012.

5. B. Escoffier, L. Gourvès, N.K. Thang, F. Pascual, and O. Spanjaard. Strategy-proof mechanisms for Facility Location games with many facilities. In *Proc. of the 2nd International Conference on Algorithmic Decision Theory (ADT '11)*, LNAI 6992, pp. 67-81, 2011.

6. M. Feldman and Y. Wilf. Strategyproof Facility Location and the least squares objective *Proc. of the 14th ACM Conference on Electronic Commerce (EC '13)*, pp. 873-890, 2013.

7. D. Fotakis and C. Tzamos. On the power of deterministic mechanisms for Facility Location games. *ACM Transactions on Economics and Computation*, 4(2), 15:1-37, 2014.

8. D. Fotakis and C. Tzamos. Winner-imposing strategyproof mechanisms for multiple Facility Location games. *Theoretical Computer Science*, 472:90-103, 2013.

9. E. Koutsoupias. Scheduling without payments. In *Proc. of the 4th International Symposium on Algorithmic Game Theory (SAGT '11)*, volume 6982 of *LNCS*, pp. 143-153, 2011.

10. P. Lu, X. Sun, Y. Wang, and Z.A. Zhu. Asymptotically Optimal Strategy-Proof Mechanisms for Two-Facility Games. In *Proc. of the 11th ACM Conference on Electronic Commerce (EC '10)*, pp. 315-324, 2010.

11. P. Lu, Y. Wang, and Y. Zhou. Tighter bounds for facility games. In *Proc. of the 5th Workshop on Internet and Network Economics (WINE '09)*, LNCS 5929, pp. 137-148, 2009.

12. P.B. Mirchandani and R.L. Francis. *Discrete Location Theory*. Wiley, 1990.

13. E. Miyagawa. Locating libraries on a street. *Social Choice and Welfare*, 18:527-541, 2001.

14. H. Moulin. On strategy-proofness and single-peakedness. *Public Choice*, 35:437-455, 1980.

15. N. Nisan, T. Roughgarden, É. Tardos, and V. Vazirani. *Algorithmic Game Theory*. Cambridge University Press, 2007.

16. K. Nissim, R. Smorodinsky, and M. Tennenholtz. Approximately optimal mechanism design via Differential Privacy. In *Proc. of the 3rd Conference on Innovations in Theoretical Computer Science (ITCS '12)*, pp. 203-213, 2012.

17. A.D. Procaccia and M. Tennenholtz. Approximate mechanism design without money. In *Proc. of the 10th ACM Conference on Electronic Commerce (EC '09)*, pp. 177-186, 2009.

18. J. Schummer and R.V. Vohra. Strategyproof location on a network. *Journal of Economic Theory*, 104:405-428, 2002.

19. Y. Sprumont. Strategyproof collective choice in economic and political environments. *The Canadian Journal of Economics*, 28(1):68-108, 1995.

20. X. Sui, C. Boutilier, and T. Sandholm. Analysis and optimization of multi-dimensional percentile mechanisms. In *Proc. of 4th International Workshop on Computational Social Choice (COMSOC '12)*, 2012.