
Procurement Auctions with Predictions: Improved Frugality for Facility Location

Eric Balkanski
Columbia University
eb3224@columbia.edu

Nicholas DeFilippis
New York University
nad9961@nyu.edu

Vasilis Gkatzelis
Drexel University,
Archimedes, Athena R.C., Greece
gkatz@drexel.edu

Xizhi Tan*
Stanford University
xizhi@stanford.edu

Abstract

We study the problem of designing procurement auctions for the strategic uncapacitated facility location problem: a company needs to procure a set of facility locations in order to serve its customers and each facility location is owned by a strategic agent. Each owner has a private cost for providing access to their facility (e.g., renting it or selling it to the company) and needs to be compensated accordingly. The goal is to design truthful auctions that decide which facilities the company should procure and how much to pay the corresponding owners, aiming to minimize the total cost, i.e., the monetary cost paid to the owners and the connection cost suffered by the customers (their distance to the nearest facility). We evaluate the performance of these auctions using the *frugality ratio*.

We first analyze the performance of the classic VCG auction in this context and prove that its frugality ratio is exactly 3. We then leverage the learning-augmented framework and design auctions that are augmented with predictions regarding the owners' private costs. Specifically, we propose a family of learning-augmented auctions that achieve significant payment reductions when the predictions are accurate, leading to much better frugality ratios. At the same time, we demonstrate that these auctions remain robust even if the predictions are arbitrarily inaccurate, and maintain reasonable frugality ratios even under adversarially chosen predictions. We finally provide a family of "error-tolerant" auctions that maintain improved frugality ratios even if the predictions are only approximately accurate, and we provide upper bounds on their frugality ratio as a function of the prediction error.

1 Introduction

When government agencies, retail chains, or banking institutions need to open new facilities or branches to better serve their customers, they face a challenging optimization problem: they want customers to be as close as possible to a facility, but opening new facilities can be very costly, and the cost can vary significantly by location. Deciding which locations to open therefore requires balancing the customers' *connection costs* (their distance to the nearest facility) against the facilities' *opening costs*. This optimization problem is known as the *uncapacitated facility location* (UFL) problem and it has received a lot of attention in prior work (e.g., [14, 21, 29, 36, 36, 40, 47]).

However, the vast majority of this prior work assumes that the designer has full information regarding both the connection and the opening costs, which is often unrealistic in practice. For example, each location may be owned by a different strategic agent who would need to be appropriately compensated

*This work was done while the author was a PhD student at Drexel University

for the agency to open a facility there. Negotiating this compensation (i.e., the opening cost) is itself a strategic problem: the owner wishes to maximize it, while the agency wishes to minimize it. This additional obstacle is captured by the *strategic* UFL problem, where the opening cost of each location is private information held by the agent who own it [3, 19, 35, 48].

From the agency’s perspective, rather than negotiating compensation separately with each owner, a more effective approach is to run a procurement auction. This encourages competition, leading to more efficient and cost-effective outcomes, which is one primary reason why the US government routinely uses procurement auctions [44]. To prevent manipulation, prior work on strategic UFL has focused on designing *truthful* auctions that incentivize agents to reveal their true opening costs [3, 19, 35, 48]. Our work continues in this direction.

Evaluating the performance of a procurement auction is complicated by the lack of a suitable benchmark. If there is little competition, the agency may have to pay higher compensation, whereas greater competition can reduce costs. The *frugality* literature [4, 48] addresses this by proposing the cost of the “second-best” solution as a benchmark (see Section 2). When this second-best cost is much higher than the optimum, the instance resembles a monopoly, and the agency must pay more. Conversely, if the second-best is close to the optimum, a well-designed auction can leverage competition for better outcomes. The *frugality ratio* quantifies how closely the auction’s cost approaches this benchmark, and our main technical results are new frugality ratio bounds.

Despite its appeal, most prior work in frugal mechanism design is restricted to the adversarial framework, analyzing the auctions’ performance under worst-case instances and assuming they have no information regarding the private costs. While this provides robust guarantees, it can be overly pessimistic, especially in practice, where agencies often have access to predictions regarding opening costs, derived from historical data, expert estimates, or data-driven models. This naturally raises the question: can we design mechanisms that exploit such predictions to improve frugality, while still remaining robust if the predictions turn out to be inaccurate?

Motivated by this, our work leverages the *learning-augmented framework*, which has recently spurred significant research on “algorithms with predictions” [37, 42], and more recently, on the design of truthful auctions and mechanisms augmented with predictions [1, 49]. In this framework, the designer is given an unreliable prediction (potentially from machine learning or historical data), and the goal is to achieve improved guarantees when the prediction is accurate (*consistency*), while maintaining strong guarantees even when the prediction is arbitrarily inaccurate (*robustness*).

1.1 Our Results

Our first result is a tight analysis of the frugality ratio of the classic truthful VCG auction (without predictions) for the strategic UFL problem. Prior to this result, the best known upper bound was 4 by Talwar [48], and there were no known lower bounds. Using a tighter analysis and a matching lower bound, we prove that its frugality ratio is exactly 3.

Then, rather than assuming that the auction has no information regarding the private opening cost o_ℓ at each location ℓ , we consider *learning-augmented auctions* that are provided with a prediction $\hat{o}_\ell$ regarding this cost. Crucially, this prediction is unreliable and can be arbitrarily inaccurate.

Our second result is a new truthful procurement auction that takes a parameter $\epsilon \in (0, 2]$ as input and guarantees a frugality ratio of $1 + \epsilon$ whenever the prediction is accurate (the *consistency* guarantee), while simultaneously guaranteeing a frugality ratio of at most $\max(5, 3 + \frac{2}{\epsilon})$, irrespective of how inaccurate the prediction may be (the *robustness* guarantee). Note that choosing a small constant ϵ allows us to guarantee a near-optimal frugality (arbitrarily close to 1) when the prediction is accurate, while simultaneously guaranteeing a constant frugality even for adversarially chosen predictions. To achieve this guarantee, our learning-augmented auction uses the VCG mechanism on input that is scaled using the predictions. Specifically, if the opening cost reported by the owner at some location ℓ exceeds the prediction $\hat{o}_\ell$, then it is scaled up even higher, reducing their chance of being selected. When the prediction is accurate, this limits the agents’ ability to manipulate, leading to reduced costs.

For our third result, we design a learning-augmented procurement auction that is “error tolerant.” We first define the prediction error as $\eta = \max_{\ell \in L} \max(\hat{o}_\ell/o_\ell, o_\ell/\hat{o}_\ell)$, i.e., the largest ratio between the predicted opening cost and the actual opening cost. Our auction is then provided with an error tolerance parameter, $\lambda > 1$, as input (as well as the parameter $\epsilon \in (0, 2]$), and as long as the error η of

the prediction is at most λ , it guarantees a frugality ratio of $\eta(1 + \lambda) + 2\epsilon$. Even if the error exceeds the error tolerance threshold, i.e., $\eta > \lambda$, the frugality ratio is at most $\max\{2\lambda^4 + 3\lambda^2, 3 + \frac{2}{\epsilon}\}$.

1.2 Related Work

The strategic UFL problem was first studied by Archer and Tardos [3], who analyzed the structure of truthful mechanisms for this problem. The frugality of truthful mechanisms for the strategic UFL problem was studied by Talwar [48], who showed that the VCG auction has a frugality ratio of at most 4. Chen et al. [19] considered different social cost objectives for the special case where the facility locations coincide with the locations of the agents (i.e., each agent has a “dual-role” as a customer or facility operator). Chen et al. [19] and Li et al. [35] also considered the budget-feasible version of this problem, where the overall payment cannot exceed a budget. For related work on the non-strategic version of the UFL problem, see Appendix A.

The study of frugal mechanism design is motivated by the challenge of minimizing unnecessary overpayment while preserving truthfulness in procurement auctions and combinatorial team selection problems. This perspective has been applied across a variety of classic optimization settings. In network design, Archer and Tardos [4] and Talwar [48] initiated the investigation for path auctions, demonstrating that VCG mechanisms can incur large overpayments and establishing tight frugality bounds under the frugal solution benchmark. For coverage problems such as vertex cover and set cover, Elkind et al. [24] and subsequent works developed truthful mechanisms with frugality ratios that depend on structural parameters like graph degree or spectral properties, and also provided matching lower bounds in several cases. Frugality in k -flow and cut problems has also been explored, with mechanisms achieving constant-competitive frugality ratios relative to various benchmarks, including both disjoint-alternative and equilibrium-based definitions [17, 33, 34]. In some settings, such as matroid and spanning tree auctions, VCG mechanisms are provably frugal, achieving optimal or near-optimal frugality ratios [33, 48].

The design of learning-augmented mechanisms for settings involving strategic agents was initiated by Agrawal et al. [1] and Xu and Lu [49]. This line of work spans strategic facility location [1, 8, 12, 18, 31, 46, 49], strategic scheduling [7, 20, 49], auction design [9, 15, 16, 28, 38, 41, 49], bicriteria mechanism design for welfare and revenue trade-offs [6], graph problems with private input [22], distortion [13, 25], and equilibrium analysis [27, 32]. The work most closely related to this paper is [49], which studied frugal mechanism design with predictions in the context of path auctions. See Appendix A for a more extensive discussion on the line of work on algorithms with predictions.

2 Preliminaries

In the *Uncapacitated Facility Location (UFL) problem*, there is a set U of users and a set L of facilities. Each facility $\ell \in L$ has an opening cost o_ℓ . Each user $u \in U$ and facility $\ell \in L$ have a connection cost $d(u, \ell)$. The connection costs are assumed to form a metric space, i.e., $d(x, y)$ is defined for any $x, y \in U \cup L$ and satisfies $d(x, x) = 0$, $d(x, y) \geq 0$ (non-negativity), $d(x, y) = d(y, x)$ (symmetry), and $d(x, z) \leq d(x, y) + d(y, z)$ (triangle inequality) for all $x, y, z \in U \cup L$. The cost of connecting a user $u \in U$ to a set of facilities $S \subseteq L$ is $\min_{\ell \in S} d(u, \ell)$, i.e., it is its distance to its closest facility in S . Given facilities $S \subseteq L$, we say that a user u is assigned to facility ℓ if $\ell = \arg \min_{\ell \in S} d(u, \ell)$. The total connection cost incurred by a set of facilities $S \subseteq L$ is

$$d(U, S) = \sum_{u \in U} \min_{\ell \in S} d(u, \ell),$$

and its total cost $c(S)$ is the sum of its opening cost and total connection cost, i.e.,

$$c(S) = \sum_{\ell \in S} o_\ell + d(U, S).$$

The goal is to open a set of facilities $S \subseteq L$ that minimizes the total cost. Given U, d , and $\mathbf{o} = (o_\ell)_{\ell \in L}$, the optimal facility set is $\text{OPT}(U, \mathbf{o}, d) = \arg \min_{S \subseteq L} c(S)$.

Strategic UFL. In the strategic version of the UFL problem, each facility $\ell \in L$ is owned by a strategic agent and its opening cost o_ℓ is private (the connection costs are public information). An auction $\mathcal{M}(U, \mathbf{b}, d)$ for strategic UFL takes as input the set of users U and the connection costs d ,

and it asks the owner of each location $\ell \in L$ to report an opening cost b_ℓ (which we refer to as a bid), leading to a bid vector $\mathbf{b} = (b_\ell)_{\ell \in L}$. For simplicity, we write $\mathcal{M}(\mathbf{b})$ and $\text{OPT}(\mathbf{o})$ when U and d are clear from context. We also use $\mathbf{b}_{-\ell}$ to refer to the vector of all bids excluding b_ℓ . We say that a facility ℓ misreports if its reported cost is not its true cost, i.e., $b_\ell \neq o_\ell$. The output $(S, \mathbf{p})$ of an auction consists of a set $S \subseteq L$ of facilities to open and a payment p_ℓ to each facility $\ell \in S$ for opening. The utility of a facility from output $(S, \mathbf{p})$ is

$$u_\ell(S, \mathbf{p}) = \begin{cases} p_\ell - o_\ell & \text{if } \ell \in S \\ 0 & \text{otherwise} \end{cases}.$$

An auction is *truthful* if it is a dominant strategy for every facility ℓ to report its true opening cost, i.e., $b_\ell = o_\ell$, irrespective of what the other facilities report. That is, for every $o_\ell, b_\ell, \mathbf{b}_{-\ell}, U, d$,

$$u_\ell(\mathcal{M}(o_\ell, \mathbf{b}_{-\ell})) \geq u_\ell(\mathcal{M}(b_\ell, \mathbf{b}_{-\ell})).$$

An auction $\mathcal{M}$ for UFL is *monotone* if for any facility $\ell \in L$ and any bid profile $\mathbf{b}_{-\ell}$ of the other facilities, the probability that $\ell \in S$, where S is the facilities chosen by $\mathcal{M}(b_\ell, \mathbf{b}_{-\ell})$, is non-increasing in its bid b_ℓ . In single-parameter environments such as strategic UFL, Myerson's Lemma characterizes truthful auctions:

Lemma 2.1 (Myerson's Lemma). *An auction is truthful if and only if it is monotone. For any monotone auction, there exists a unique payment rule that ensures truthfulness, which can be computed explicitly.*

Frugality and Efficiency. The total cost incurred by an auction for outputting $(S, \mathbf{p})$ is the sum of the payments and the total connection cost, i.e.,

$$p(S, \mathbf{p}) = \sum_{\ell \in S} p_\ell + d(U, S).$$

The *frugal facility set* is the second-best solution, i.e.,

$$F(U, \mathbf{o}, d) = \arg \min_{S \subseteq L \setminus \text{OPT}(U, \mathbf{o}, d)} c(S).$$

The efficiency of an auction is evaluated by its *frugality ratio*, which is the worst-case ratio between its total cost and the cost of the frugal solution:

$$\text{frugality}(\mathcal{M}) = \max_{U, \mathbf{o}, d} \frac{p(\mathcal{M}(U, \mathbf{o}, d))}{c(F(U, \mathbf{o}, d))}.$$

Learning-Augmented Framework. In the *learning-augmented setting*, the auction is given predictions $\hat{\mathbf{o}}$ about the opening costs $\mathbf{o}$. We evaluate the performance of $\mathcal{M}$ using two metrics.

Consistency: The frugality ratio when the predictions are accurate ($\hat{\mathbf{o}} = \mathbf{o}$):

$$\text{consistency}(\mathcal{M}) = \max_{U, \mathbf{o}, d} \frac{p(\mathcal{M}(U, \mathbf{o}, d, \hat{\mathbf{o}} = \mathbf{o}))}{c(F(U, \mathbf{o}, d))}.$$

Robustness: The frugality ratio when the predictions can be arbitrarily wrong:

$$\text{robustness}(\mathcal{M}) = \max_{U, \mathbf{o}, d, \hat{\mathbf{o}}} \frac{p(\mathcal{M}(U, \mathbf{o}, d, \hat{\mathbf{o}}))}{c(F(U, \mathbf{o}))}.$$

3 Tight Frugality Bounds for the Vickrey-Clarke-Groves Auction

Our first main result establishes an improved frugality ratio of 3 for the VCG auction, which we start by describing in the context of the strategic UFL problem.

The VCG auction. Given an instance $(U, \mathbf{o}, d)$ of UFL, the set of facilities opened by the VCG auction is the optimal facility set $\text{OPT}(\mathbf{o})$. The VCG payment, also called threshold payment, to each opened facility $\ell \in \text{OPT}(\mathbf{o})$ is

$$p_\ell = c(\text{OPT}(\infty, o_{-\ell})) - c(\text{OPT}(0, \mathbf{o}_{-\ell})),$$

i.e., it is the difference between the total cost of the optimal solution when facility ℓ is infeasible ($o_\ell = \infty$) and when ℓ is free ($o_\ell = 0$). This payment can be alternatively defined as

$$p_\ell = \max\{b \geq 0 : \ell \in \text{OPT}(b, \mathbf{o}_{-\ell})\},$$

i.e., the largest bid that facility ℓ can declare and remain in the optimal set OPT selected by VCG.

Theorem 3.1. *The frugality ratio of the VCG auction for the metric uncapacitated facility location problem is exactly 3.*

We first argue an upper bound of 3. At a high level, the proof bounds the total cost incurred by the VCG auction by (1) bounding the payments by twice the total cost of the frugal solution and (2) observing that the total connection cost of the optimal solution is bounded by its own total cost, which is in turn bounded by the total cost of the frugal solution. Combining these two facts immediately yields a 3-approximation. The payment bound is achieved by a “rerouting” argument: for each facility in the optimal set, we show how to reassign its users to other facilities, either within the optimal set itself or to those in the frugal solution, thereby upper-bounding its threshold payment by the corresponding rerouting cost.

Proof of Theorem 3.1. Let OPT and F be the optimal and frugal solutions. We denote by $U_\ell \subseteq U$ the users assigned to facility ℓ according to OPT . We first note that for any set of facilities $O_\ell \subseteq L \setminus \{\ell\}$ that does not contain facility ℓ , we must have

$$p_\ell + \sum_{u \in U_\ell} d(u, \ell) \leq \sum_{f \in O_\ell \setminus \text{OPT}} o_f + \sum_{u \in U_\ell} \min_{f' \in O_\ell} d(u, f'), \quad (1)$$

otherwise O_ℓ would have a smaller total cost than OPT when facility ℓ reports opening cost p_ℓ , which would imply $p_\ell > \max\{b \geq 0 : \ell \in \text{OPT}(b, \mathbf{o}_{-\ell})\}$ and would contradict the alternative definition of VCG.

The central part of the proof consists of defining some alternative solution O_ℓ and to reassign and reroute the users in U_ℓ to facilities in O_ℓ to bound $\sum_{u \in U_\ell} \min_{f' \in O_\ell} d(u, f')$. To define O_ℓ and this reassignment, we first need to introduce some notation. For each pair $(\ell \in \text{OPT}, f \in F)$, we let $U_{\ell, f} \subseteq U$ be the set of users assigned to ℓ in the optimal solution OPT and to f in the frugal solution F . For each $f \in F$, let $\text{OPT}_f = \{\ell \in \text{OPT} : |U_{\ell, f}| > 0\}$ and $x_f(\ell)$ be the ranking of $\ell \in \text{OPT}_f$ among all other facilities in OPT_f in terms of the size of $|U_{\ell, f}|$, where ties are broken arbitrarily but consistently. This definition implies that $x_f^{-1}(j) \in \text{OPT}_f$ is the facility ℓ with the j^{th} smallest non-negative $|U_{\ell, f}|$ and we thus have $0 < |U_{x_f^{-1}(1), f}| \leq \dots \leq |U_{x_f^{-1}(|\text{OPT}_f|), f}|$.

We define the alternate solution to be $O_\ell = \text{OPT} \setminus \{\ell\} \cup \{f \in F : x_f(\ell) = |\text{OPT}_f|\}$. In this solution, facility ℓ is removed from OPT and the facilities $f \in F$ such that $|U_{\ell, f}|$ is largest among $\{|U_{\ell', f}|\}_{\ell' \in \text{OPT}}$ are added. We then define a mapping $\pi_f : \text{OPT}_f \rightarrow \text{OPT}_f \cup \{f\}$ that reassigns users in $U_{\ell, f}$ that are assigned to facility ℓ according to OPT to a new facility in $\text{OPT}_f \cup \{f\}$. This reassignment is defined as

$$\pi_f(\ell) = \begin{cases} x_f^{-1}(x_f(\ell) + 1) & \text{if } x_f(\ell) < |\text{OPT}_f| \\ f & \text{if } x_f(\ell) = |\text{OPT}_f| \end{cases}.$$

In words, $\pi_f(\ell)$ maps facility ℓ to the facility ranked immediately after ℓ with respect to f ; if ℓ is ranked last, then $\pi_f(\ell) = f$. Since $\cup_{f: \ell \in \text{OPT}_f} U_{\ell, f} = U_\ell$, we have that

$$\sum_{u \in U_\ell} \min_{f' \in O_\ell} d(u, f') \leq \sum_{f: \ell \in \text{OPT}_f} \min_{f' \in O_\ell} d(U_{\ell, f}, f') \leq \sum_{f: \ell \in \text{OPT}_f} d(U_{\ell, f}, \pi_f(\ell)). \quad (2)$$

Next, if $x_f(\ell) < |\text{OPT}_f|$, then $|U_{\ell, f}| \leq |U_{\pi_f(\ell), f}|$ by the definition of π_f . Thus, combined with the triangle inequality, we get that if $x_f(\ell) < |\text{OPT}_f|$, then

$$d(U_{\ell, f}, \pi_f(\ell)) \leq \sum_{f: \ell \in \text{OPT}_f} d(U_{\ell, f}, f) + d(f, U_{\pi_f(\ell), f}) + d(U_{\pi_f(\ell), f}, \pi_f(\ell)). \quad (3)$$

Informally, this last inequality corresponds to rerouting the users in $U_{\ell, f}$ to $\pi_f(\ell)$ by going through f and then $U_{\pi_f(\ell), f}$. This rerouting is the key component of this proof, see Figure 1 for an illustration.

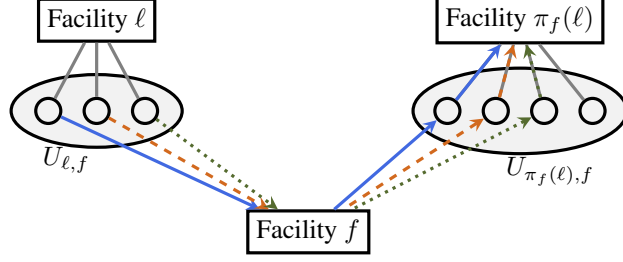


Figure 1: Illustration of how users in $U_{\ell,f}$ are rerouted to facility $\pi_f(\ell)$. Each color (and line pattern) denotes the path taken by a distinct user $u \in U_{\ell,f}$. The total connection-cost along each set of colored edges gives an upper bound on the payment to facility ℓ for serving that subset of users.

We obtain that

$$\begin{aligned}
p(\text{OPT}) &= \sum_{\ell \in \text{OPT}} \left(p_\ell + \sum_{u \in U_\ell} d(u, \ell) \right) && \text{by def. of } p \\
&\leq \sum_{\ell \in \text{OPT}} \left(\sum_{f: x_f(\ell) = |\text{OPT}_f|} o_f + \sum_{u \in U_\ell} \min_{f' \in O_\ell} d(u, f') \right) && \text{by (1)} \\
&\leq \sum_{\ell \in \text{OPT}} \left(\sum_{f: x_f(\ell) = |\text{OPT}_f|} o_f + \sum_{f: \ell \in \text{OPT}_f} d(U_{\ell,f}, \pi_f(\ell)) \right) && \text{by (2)} \\
&= \sum_{f \in F} \left(o_f + \sum_{\ell \in \text{OPT}_f} d(U_{\ell,f}, \pi_f(\ell)) \right) \\
&= \sum_{f \in F} \left(o_f + d(U_{x_f^{-1}(|\text{OPT}_f|),f}, f) + \sum_{j=1}^{|\text{OPT}_f|-1} d(U_{x_f^{-1}(j),f}, x_f^{-1}(j+1)) \right) \\
&\leq \sum_{f \in F} \left(o_f + d(U_{x_f^{-1}(|\text{OPT}_f|),f}, f) \right. \\
&\quad \left. + \sum_{j=1}^{|\text{OPT}_f|-1} \left(d(U_{x_f^{-1}(j),f}, f) + d(f, U_{x_f^{-1}(j+1),f}) \right. \right. \\
&\quad \left. \left. + d(U_{x_f^{-1}(j+1),f}, x_f^{-1}(j+1)) \right) \right) && \text{by (3)} \\
&\leq \sum_{f \in F} \left(o_f + \sum_{\ell \in \text{OPT}_f} (d(U_{\ell,f}, f) + d(f, U_{\ell,f}) + d(U_{\ell,f}, \ell)) \right) \\
&= \left(\sum_{f \in F} o_f + 2 \sum_{f \in F} \sum_{\ell \in \text{OPT}_f} d(U_{\ell,f}, f) \right) + \sum_{\ell \in \text{OPT}} \sum_{f: \ell \in \text{OPT}_f} d(U_{\ell,f}, \ell) \\
&\leq 2 \cdot c(F) + c(\text{OPT}) \\
&\leq 3 \cdot c(F).
\end{aligned}$$

We show that this upper bound of 3 is tight for VCG by providing a simple tree-metric instance. Due to space limitations, we defer its analysis to Appendix B. $\square$

4 Learning-Augmented Frugal Auction

We now consider the strategic UFL problem within the learning-augmented framework, where the auction has access to potentially inaccurate predictions about the true opening costs of the facilities. Specifically, for each facility $\ell \in L$, a prediction $\hat{o}_\ell$ is provided for its true opening cost o_ℓ . Our objective is to leverage this unreliable information to design truthful auctions that minimize the frugality ratio while remaining robust against inaccurate predictions.

To achieve this, we propose PREDICTEDLIMITS, described in Auction 1. This auction first identifies the optimal solution based on the predicted costs, denoted by OPT . It then modifies the cost function by scaling the actual opening costs of facilities in OPT when their true costs exceed the predicted opening costs. Finally, the auction selects the solution that minimizes this scaled cost function. We note that,

Auction 1: PREDICTEDLIMITS

Input: Set of clients U , set of facilities L , parameter ϵ , predicted opening costs $\hat{o}_\ell$ for all $\ell \in L$

Output: Selected facility subset S^* and threshold payments for each facility $\ell \in S^*$

Compute $\text{OPT} \leftarrow \arg \min_{S \subseteq L} (d(U, S) + \sum_{\ell \in S} \hat{o}_\ell)$;

Define the modified cost function for any subset $S \subseteq L$:

$$o'_\ell(S) \leftarrow \begin{cases} \frac{2}{\epsilon} \cdot o_\ell, & \text{if } S = \text{OPT and } o_\ell > \hat{o}_\ell \\ o_\ell, & \text{otherwise} \end{cases}$$

Compute:

$$S^* \leftarrow \arg \min_{S \subseteq L} \left(d(U, S) + \sum_{\ell \in S} o'_\ell(S) \right)$$

return S^* and threshold payments for each $\ell \in S^*$;

to achieve the claimed results, our auction only needs access to the membership of the predicted optimal solution OPT and to the opening costs of those facilities. We show that PREDICTEDLIMITS can achieve near-optimal consistency (arbitrarily close to 1) while still maintaining competitive robustness.

Theorem 4.1. *Given any $\epsilon \in [0, 2]$ as input, PREDICTEDLIMITS is truthful, $(1 + \epsilon)$ -consistent, and $\max(5, 3 + \frac{2}{\epsilon})$ -robust.*

At first glance, this auction may seem counterintuitive and differs from existing mechanisms that also use predictions to scale costs (e.g., [7, 49]). Instead of scaling down the opening costs of facilities in the predicted optimal solution, a set that would normally be favored for minimizing the overall social cost, the auction scales these costs upwards, making the predicted optimal solution less appealing. The primary rationale for this upward scaling is to restrict the potential overbidding the bidder can claim while remaining in the optimal solution, which leads to a high threshold price. By amplifying any over-reported opening costs over the predicted ones, the auction gains tighter control over facility payments, which can lead to a lower frugality ratio when the predictions are accurate.

Importantly, the scaling applied by our auction is *solution-dependent* rather than facility-dependent. Specifically, the opening cost scaling is activated only when evaluating the predicted optimal solution OPT as a complete set. For any other subset, the opening costs of individual facilities remain unscaled. This distinction is crucial for our robustness analysis, as it ensures that the adverse effects of inaccurate predictions are confined to the evaluation of the single predicted optimal solution, leaving alternative solutions unaffected. We defer the truthfulness proof to Appendix C.1. We first provide the consistency analysis of Auction PREDICTEDLIMITS. We then provide the robustness guarantee of the auction, which holds regardless of the quality of the predictions.

Lemma 4.2. *Given any $\epsilon \in [0, 2]$ as input, PREDICTEDLIMITS achieves a consistency of $1 + \epsilon$.*

Proof. Recall that consistency analysis assumes the predictions are correct, i.e., $\hat{o}_\ell = o_\ell$ for all facilities $\ell \in L$. By the definition of Auction 1, if predictions are accurate, the condition $o_\ell > \hat{o}_\ell$ for $\ell \in \text{OPT}$ (which equals OPT here) is never met. Therefore, no costs are scaled, and the auction selects the true optimal solution, i.e., $S^* = \text{OPT}$. We adopt the same definitions of OPT_f and $\pi_f(\ell)$ as in the analysis of VCG (proof of Theorem 3.1).

We now provide an upper bound for the payment p_ℓ to the agent controlling facility $\ell \in \text{OPT}$. Similarly to VCG, we will upper-bound the payment by considering the cost of rerouting. The difference is that when the reported bid exceeds $\hat{o}_\ell$, the scaling would apply. Formally, for any set of facilities $O_\ell \subseteq L \setminus \{\ell\}$ not containing ℓ :

$$\frac{2}{\epsilon} p_\ell + \sum_{u \in U_\ell} d(u, \ell) \leq \sum_{f' \in O_\ell \setminus \text{OPT}} o_{f'} + \sum_{u \in U_\ell} \min_{f' \in O_\ell} d(u, f'). \quad (4)$$

Note that the scaling affects only the opening costs of the facilities in $\hat{\text{OPT}}$. Hence the right-hand side of (4) remains unchanged for all (ℓ, f) pairs relative to the VCG analysis, as it represents cost of alternative solutions not involving ℓ 's potentially scaled cost. Therefore, the bounds from Equations (2) and (3) continue to hold. Following the same steps as in the VCG proof (Theorem 3.1), we obtain:

$$\begin{aligned} \sum_{\ell \in \text{OPT}} \left(\frac{2}{\epsilon} p_\ell + \sum_{u \in U_\ell} d(u, \ell) \right) &\leq \left(\sum_{f \in F} o_f + 2 \sum_{f \in F} \sum_{\ell \in \text{OPT}_f} d(U_{\ell, f}, f) \right) + \sum_{\ell \in \text{OPT}} \sum_{f: \ell \in \text{OPT}_f} d(U_{\ell, f}, \ell) \\ &\Rightarrow \frac{2}{\epsilon} \sum_{\ell \in \text{OPT}} p_\ell \leq 2 \sum_{f \in F} o_f + 2 \sum_{f \in F} \sum_{\ell \in \text{OPT}_f} d(U_{\ell, f}, f) \Rightarrow \sum_{\ell \in \text{OPT}} p_\ell \leq \epsilon \cdot c(F). \end{aligned}$$

Together with the fact that the connection cost of the optimal solution is at most $c(F)$, we conclude that the total auction cost $\sum_{\ell \in \text{OPT}} p_\ell + d(U, \text{OPT})$ is at most $(1 + \epsilon) c(F)$. $\square$

We now move on to the robustness analysis. We begin with the simpler case in which the auction still outputs the optimal solution of the instance, i.e., $S^* = \text{OPT}$. All missing proofs of the section can be found in Appendix C.2.

Lemma 4.3. *Given any instance $(U, \mathbf{o}, d)$ and any $\epsilon \leq 2$, let OPT and F be the optimal solution and frugal solution, respectively. If auction 1 outputs $S^* = \text{OPT}$, then we have:*

$$\sum_{\ell \in S^*} p_\ell + d(U, S^*) \leq \left(1 + \frac{2}{\epsilon} \right) \sum_{\ell \in F} o_\ell + 3 d(U, F)$$

The proof is deferred to Appendix C.2.1. We now address the case where scaling alters the output set, i.e., $S^* \neq \text{OPT}$. This case is particularly challenging because S^* may contain facilities belonging to both the frugal solution and the optimal solution, making it difficult to bound the payment. To tackle this, we need a different accounting argument than the ones used for the VCG and consistency analyses. Specifically, we will analyze payments for facilities in the output solution S^* differently, based on whether they are part of the frugal solution F . We begin by establishing a bound for payments to non-frugal facilities in S^* (i.e., those in $S^* \setminus F$) with the following lemma.

Lemma 4.4. *Given any instance $(U, \mathbf{o}, d)$, and any $\epsilon \leq 2$, let OPT and F be the optimal solution and frugal solution, respectively. Let $S^* \neq \text{OPT}$ be the output of Auction 1, and let $S = S^* \setminus F$. Then:*

$$\sum_{\ell \in S} p_\ell \leq \sum_{\ell \in F} o_\ell + 2d(U, F).$$

The proof is deferred to Appendix C.2.2. Then, we bound the payment to the output solution that is also part of the frugal solution.

Lemma 4.5. *Given any instance $(U, \mathbf{o}, d)$, and any $\epsilon \leq 2$, let OPT and F be the optimal solution and frugal solution, respectively. Let $S^* \neq \text{OPT}$ be the output of Auction 1, and $S^f = S^* \cap F$. Then:*

$$\sum_{\ell \in S^f} p_\ell \leq \max \left(2, \frac{2}{\epsilon} \right) \cdot \left[\sum_{\ell \in F} o_\ell + d(U, F) \right].$$

Proof. The approach considers rerouting users currently assigned to a facility $f \in S^f$ back to their corresponding facility ℓ in the true optimal solution OPT . This facility $j \in \text{OPT}$ might not be included in the output set S^* , and if $\ell \in \text{OPT} = \hat{\text{OPT}}$, its opening cost o_ℓ might be effectively scaled

upwards to o'_j in the auction's payment calculation if it was under-predicted ($o_\ell > \hat{o}_\ell$). The analysis mirrors the VCG structure but with the roles of the chosen solution (S^f) and the alternative (OPT) flipped. Define U_ℓ , $U_{f,\ell}$, and $\pi_\ell(f)$ as in the proof of Theorem 3.1, replacing OPT with S^f . Define $S_\ell^f = \{\ell \in S^f : |U_{f,\ell}| > 0\}$.

Case 1: The optimal facility $\ell \notin S^*$. By the same rerouting argument as VCG, inequalities (2) and (3) hold with OPT replaced by S^f and the opening cost replaced by the scaled opening cost.

Case 2: The optimal facility $\ell \in S^*$. Here, users $U_{f,\ell}$ can be rerouted directly to ℓ within the output set S^* , so the cost associated with rerouting users $U_{f,\ell}$, denoted obeys

$$p_\ell + \sum_{u \in U_\ell} d(u, \ell) \leq d(U_{f,\ell}, \ell).$$

This bound is better than inequalities (2) and (3) from Case 1.

Summing over all agents $f \in S^f$ and assuming the worst-case where every cost gets scaled we get:

$$\begin{aligned} \sum_{\ell \in S^f} \left(p_\ell + \sum_{u \in U_\ell} d(u, \ell) \right) &= \sum_{\ell \in \text{OPT}} \sum_{f \in S_\ell^f} \left(p_\ell + \sum_{u \in U_\ell} d(u, \ell) \right) \\ &\leq \sum_{\ell \in \text{OPT}} \left(\frac{2}{\epsilon} o_\ell + 2 \sum_{f \in S_\ell^f} d(U_{\ell,f}, f) \right) + \sum_{\ell \in \text{OPT}} \sum_{f: \ell \in \text{OPT}_f} d(U_{\ell,f}, \ell) \\ &\Rightarrow \sum_{\ell \in S^f} p_\ell \leq \sum_{\ell \in \text{OPT}} \frac{2}{\epsilon} o_\ell + 2 \sum_{\ell \in \text{OPT}} \sum_{f \in S_\ell^f} d(U_{\ell,f}, f) \\ &\Rightarrow \sum_{\ell \in \text{OPT}} p_\ell \leq \max \left(\frac{2}{\epsilon}, 2 \right) c(F) = \max \left(\frac{2}{\epsilon}, 2 \right) \left[\sum_{\ell \in F} o_\ell + d(U, F) \right]. \quad \square \end{aligned}$$

By combining Lemma 4.3, Lemma 4.4, and Lemma 4.5, we get the robustness of Auction PREDICTEDLIMITS. The proof is deferred to Appendix C.2.3.

Lemma 4.6. *Given any $\epsilon \in [0, 2]$ as input, PREDICTEDLIMITS achieves a robustness of $\max(5, 3 + \frac{2}{\epsilon})$.*

5 The Error-Tolerant Scaled VCG

Auction PREDICTEDLIMITS achieves $(1 + \epsilon)$ -consistency in its frugality ratio when provided with perfect predictions of opening costs, and maintains a $\max(5, 3 + \frac{2}{\epsilon})$ -robustness guarantee in adversarial scenarios. These existing bounds, however, capture only the extremes of prediction quality. In this section, we extend the auction from the previous section to achieve an improved frugality ratio, not only for accurate predictions but also for approximately accurate predictions, thereby offering a more granular performance characterization.

We denote the prediction error as $\eta \geq 1$, where η is defined as the largest ratio between the predicted opening cost and the actual opening cost, i.e.,

$$\eta = \max_{\ell \in L} \max \left(\frac{\hat{o}_\ell}{o_\ell}, \frac{o_\ell}{\hat{o}_\ell} \right).$$

The auction, ERRORTOLERANT, formally defined in Auction 2, extends PREDICTEDLIMITS by incorporating an error-tolerance parameter $\lambda \geq 0$. Recall that the original auction selects the predicted-optimal facility set $\hat{\text{OPT}}$ and, whenever a facility's true opening cost exceeds its prediction, inflates that cost by a factor of $2/\epsilon$; otherwise, it leaves opening costs unchanged.

ERRORTOLERANT is designed to preserve this behavior not only under exact predictions but also when predictions are accurate up to a factor of λ , with one key enhancement. To ensure that $\hat{\text{OPT}}$ remains the chosen set under low prediction error, we replace the overscaling threshold $\hat{o}_\ell$ with $\lambda \hat{o}_\ell$. In addition, if every facility in $\hat{\text{OPT}}$ reports a true cost below this new threshold, we apply a uniform

Auction 2: ERRORTOLERANT

Input: Clients U , facilities L , predicted costs $\hat{o}_\ell$ for all $\ell \in L$, parameter $\epsilon \leq 2$, error-tolerance $\lambda \geq 0$

Output: Chosen facilities S^* and threshold payments for each $\ell \in S^*$

$\text{OPT} \leftarrow \arg \min_{S \subseteq L} (d(U, S) + \sum_{\ell \in S} \hat{o}_\ell);$

if $\forall \ell \in \text{OPT}, o_\ell \leq \lambda \hat{o}_\ell$ **then**

$$\text{Define } c_\lambda(S) = \begin{cases} \frac{1}{\lambda^2} (d(U, S) + \sum_{\ell \in S} o_\ell), & S = \text{OPT}, \\ d(U, S) + \sum_{\ell \in S} o_\ell, & \text{otherwise.} \end{cases}$$

else

$$\text{Define for each } \ell \text{ and } S \quad o'_\ell(S) = \begin{cases} \frac{2}{\epsilon} o_\ell, & S = \text{OPT} \text{ and } o_\ell > \lambda \hat{o}_\ell, \\ o_\ell, & \text{otherwise,} \end{cases}$$

$$\text{and let } c_\lambda(S) = d(U, S) + \sum_{\ell \in S} o'_\ell(S).$$

$$S^* \leftarrow \arg \min_{S \subseteq L} c_\lambda(S),$$

return S^* and threshold payments for each $\ell \in S^*$;

down-scaling to the entire solution cost (both opening and connection costs). This extra adjustment guarantees that, when the prediction error is at most λ , the auction outputs OPT . Aside from these two modifications, ERRORTOLERANT follows PREDICTEDLIMITS exactly. It achieves an approximation guarantee of $\eta(1 + \lambda) + 2\epsilon$ when $\eta \leq \lambda$, and maintains a frugality bound of $\max(2\lambda^4 + 3\lambda^2, 3 + 2/\epsilon)$ for arbitrary error η .

Theorem 5.1. *Auction ERRORTOLERANT is truthful, and given parameters $\epsilon \in (0, 2]$ and $\lambda > 1$, it achieves the following frugality ratio, where η is the error of the prediction:*

$$\begin{cases} \eta(1 + \lambda) + 2\epsilon, & \text{if } \eta \leq \lambda, \\ \max\left\{2\lambda^4 + 3\lambda^2, 3 + \frac{2}{\epsilon}\right\}, & \text{if } \eta > \lambda. \end{cases}$$

Due to space limitations, we defer the analysis to Appendix E.

6 Conclusion and Open Problems

In this work, we studied the design of frugal procurement auctions for the strategic uncapacitated facility location problem. We first established a tight frugality ratio of 3 for the classic VCG auction. Prior to our work, the best known upper bound for VCG was 4 [48] and there was no known lower bound. We then considered the problem in the learning-augmented framework where the auction is provided with predictions regarding the costs of the facilities. We designed a novel truthful auction that achieves a frugality ratio of $1 + \epsilon$ when the predictions are accurate, while maintaining a constant-factor robustness guarantee even when the predictions are arbitrarily wrong.

Although the focus of our work is on the information-theoretic limitations of truthful auctions, we note that our proposed mechanisms can be formulated as integer programs (IPs). Please see Appendix F for the IP implementation.

Our work leaves several exciting directions open. In the classic setting, while we show that VCG's frugality ratio is exactly 3, the optimal frugality ratio achievable by a truthful auction remains unknown. An important question is whether VCG is optimal, or if another design can achieve a better frugality ratio. Thus, establishing a tight lower bound for this problem is a key next step. In the learning-augmented setting, an open problem is to determine the optimal trade-off between consistency and robustness, thereby characterizing the Pareto-optimal frontier. Furthermore, exploring more nuanced definitions of prediction error beyond magnitude—such as the number of mispredictions or the recent "mostly and approximately correct" (MAC) framework [12]—is a promising direction. Analyzing mechanisms under such models could lead to auctions that degrade more gracefully given imperfect, yet informative, predictions.

Acknowledgments and Disclosure of Funding

Eric Balkanski was supported by NSF grants CCF-2210501 and IIS-2147361. Nicholas DeFilippis was supported by an NSF Graduate Research Fellowship under Grant No. DGE-2039655. Vasilis Gkatzelis and Xizhi Tan were supported by NSF CAREER award CCF-2047907 and NSF grant CCF-2210502. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

- [1] P. Agrawal, E. Balkanski, V. Gkatzelis, T. Ou, and X. Tan. Learning-augmented mechanism design: Leveraging predictions for facility location. *Math. Oper. Res.*, 49(4):2626–2651, 2024. doi: 10.1287/MOOR.2022.0225. URL <https://doi.org/10.1287/moor.2022.0225>.
- [2] A. Antoniadis, T. Gouleakis, P. Kleer, and P. Kolev. Secretary and online matching problems with machine learned advice. *Discret. Optim.*, 48(Part 2):100778, 2023.
- [3] A. Archer and É. Tardos. Truthful mechanisms for one-parameter agents. In *42nd Annual Symposium on Foundations of Computer Science, FOCS 2001, 14-17 October 2001, Las Vegas, Nevada, USA*, pages 482–491. IEEE Computer Society, 2001.
- [4] A. Archer and É. Tardos. Frugal path mechanisms. In *Proceedings of the Thirteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 991–999. SIAM, 2002.
- [5] Y. Azar, D. Panigrahi, and N. Toudou. Online graph algorithms with predictions. *Proceedings of the Thirty-Third Annual ACM-SIAM Symposium on Discrete Algorithms*, 2022.
- [6] M.-F. Balcan, S. Prasad, and T. Sandholm. Bicriteria multidimensional mechanism design with side information. In *Proceedings of the 37th International Conference on Neural Information Processing Systems, NIPS ’23, Red Hook, NY, USA, 2024*. Curran Associates Inc.
- [7] E. Balkanski, V. Gkatzelis, and X. Tan. Strategyproof scheduling with predictions. In Y. T. Kalai, editor, *14th Innovations in Theoretical Computer Science Conference, ITCS 2023, January 10-13, 2023, MIT, Cambridge, Massachusetts, USA*, volume 251 of *LIPIcs*, pages 11:1–11:22. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2023.
- [8] E. Balkanski, V. Gkatzelis, and G. Shahkarami. Randomized strategic facility location with predictions. In A. Globersons, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. M. Tomczak, and C. Zhang, editors, *Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024*, 2024.
- [9] E. Balkanski, V. Gkatzelis, X. Tan, and C. Zhu. Online mechanism design with predictions. In D. Bergemann, R. Kleinberg, and D. Sabán, editors, *Proceedings of the 25th ACM Conference on Economics and Computation, EC 2024, New Haven, CT, USA, July 8-11, 2024*, page 1184. ACM, 2024.
- [10] É. Bamas, A. Maggiori, and O. Svensson. The primal-dual method for learning augmented algorithms. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020.
- [11] S. Banerjee, V. Gkatzelis, A. Gorokh, and B. Jin. Online nash social welfare maximization with predictions. In *Proceedings of the 2022 ACM-SIAM Symposium on Discrete Algorithms, SODA 2022*. SIAM, 2022.
- [12] Z. Barak, A. Gupta, and I. Talgam-Cohen. Mac advice for facility location mechanism design. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang, editors, *Advances in Neural Information Processing Systems*, volume 37, pages 129564–129604. Curran Associates, Inc., 2024.

- [13] B. Berger, M. Feldman, V. Gkatzelis, and X. Tan. Learning-augmented metric distortion via (p, q) -veto core. In D. Bergemann, R. Kleinberg, and D. Sabán, editors, *Proceedings of the 25th ACM Conference on Economics and Computation, EC 2024, New Haven, CT, USA, July 8-11, 2024*, page 984. ACM, 2024.
- [14] J. Byrka and K. I. 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. URL <https://doi.org/10.1137/070708901>.
- [15] I. Caragiannis and G. Kalantzis. Randomized learning-augmented auctions with revenue guarantees. *arXiv preprint arXiv:2401.13384*, 2024.
- [16] I. Caragiannis and G. Kalantzis. Mechanisms for selling an item among a strategic bidder and a profiled agent. *arXiv preprint arXiv:2502.12313*, 2025.
- [17] N. Chen, E. Elkind, N. Gravin, and S. Petrov. Frugal mechanism design via spectral techniques. In *51st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 755–764. IEEE, 2010.
- [18] Q. Chen, N. Gravin, and S. Im. Strategic facility location via predictions. *arXiv preprint arXiv:2410.07497*, 2024.
- [19] X. Chen, M. Li, C. Wang, C. Wang, M. Zhang, and Y. Zhao. Mechanisms for dual-role-facility location games: Truthfulness and approximability. *Theor. Comput. Sci.*, 932:69–83, 2022.
- [20] G. Christodoulou, A. Sgouritsa, and I. Vlachos. Mechanism design augmented with output advice. *Advances in Neural Information Processing Systems*, 37:41934–41953, 2024.
- [21] F. A. Chudak and D. B. Shmoys. Improved approximation algorithms for the uncapacitated facility location problem. *SIAM J. Comput.*, 33(1):1–25, 2003. doi: 10.1137/S0097539703405754. URL <https://doi.org/10.1137/S0097539703405754>.
- [22] R. Colini-Baldeschi, S. Klumper, G. Schäfer, and A. Tsikiris. To trust or not to trust: Assignment mechanisms with predictions in the private graph model. In *Proceedings of the 25th ACM Conference on Economics and Computation*, pages 1134–1154, 2024.
- [23] P. Dütting, S. Lattanzi, R. Paes Leme, and S. Vassilvitskii. Secretaries with advice. *Mathematics of Operations Research*, 49(2):856–879, 2024. doi: 10.1287/moor.2023.1384. URL <https://doi.org/10.1287/moor.2023.1384>.
- [24] E. Elkind, L. A. Goldberg, and P. Goldberg. Frugality ratios and improved truthful mechanisms for vertex cover auctions. In *Proceedings of the 8th ACM Conference on Electronic Commerce (EC)*, pages 336–345, 2007.
- [25] A. Filos-Ratsikas, G. Kalantzis, and A. A. Voudouris. Utilitarian distortion with predictions. In *Proceedings of the 26th ACM Conference on Economics and Computation, EC ’25*, page 254–271, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400719431.
- [26] K. Fujii and Y. Yoshida. The secretary problem with predictions. *Mathematics of Operations Research*, 49(2):1241–1262, 2024. doi: 10.1287/moor.2022.0031. URL <https://doi.org/10.1287/moor.2022.0031>.
- [27] V. Gkatzelis, K. Kollias, A. Sgouritsa, and X. Tan. Improved price of anarchy via predictions. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pages 529–557, 2022.
- [28] V. Gkatzelis, D. Schoepflin, and X. Tan. Clock auctions augmented with unreliable advice. In Y. Azar and D. Panigrahi, editors, *Proceedings of the 2025 Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2025, New Orleans, LA, USA, January 12-15, 2025*, pages 2629–2655. SIAM, 2025.

- [29] S. Guha and S. Khuller. Greedy strikes back: Improved facility location algorithms. *Journal of Algorithms*, 31(1):228–248, 1999. ISSN 0196-6774. doi: <https://doi.org/10.1006/jagm.1998.0993>. URL <https://www.sciencedirect.com/science/article/pii/S0196677498909932>.
- [30] S. Im, R. Kumar, M. M. Qaem, and M. Purohit. Online knapsack with frequency predictions. In *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pages 2733–2743, 2021.
- [31] G. Istrate and C. Bonchis. Mechanism design with predictions for obnoxious facility location. *CoRR*, abs/2212.09521, 2022.
- [32] G. Istrate, C. Bonchis, and V. Bogdan. Equilibria in multiagent online problems with predictions. *CoRR*, abs/2405.11873, 2024.
- [33] A. R. Karlin, D. Kempe, and T. Tamir. Beyond vcg: Frugality of truthful mechanisms. In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 615–624. IEEE, 2005.
- [34] D. Kempe, M. Salek, and C. Moore. Frugal and truthful mechanisms for vertex covers, flows, and cuts. In *51st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 745–754. IEEE, 2010.
- [35] M. Li, C. Wang, and M. Zhang. Budget feasible mechanisms for facility location games with strategic facilities. *Auton. Agents Multi Agent Syst.*, 36(2):35, 2022.
- [36] S. 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. URL <https://doi.org/10.1016/j.ic.2012.01.007>.
- [37] A. Lindermayr and N. Megow. Alps, 2024. URL <https://algorithms-with-predictions.github.io/>.
- [38] P. Lu, Z. Wan, and J. Zhang. Competitive auctions with imperfect predictions. In *Proceedings of the 25th ACM Conference on Economics and Computation*, pages 1155–1183, 2024.
- [39] T. Lykouris and S. Vassilvitskii. Competitive caching with machine learned advice. In *International Conference on Machine Learning*, pages 3296–3305. PMLR, 2018.
- [40] M. Mahdian, Y. Ye, and J. Zhang. Approximation algorithms for metric facility location problems. *SIAM J. Comput.*, 36(2):411–432, 2006. doi: 10.1137/S0097539703435716. URL <https://doi.org/10.1137/S0097539703435716>.
- [41] A. M. Medina and S. Vassilvitskii. Revenue optimization with approximate bid predictions. In I. Guyon, U. von Luxburg, S. Bengio, H. M. Wallach, R. Fergus, S. V. N. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 1858–1866, 2017. URL <https://proceedings.neurips.cc/paper/2017/hash/884d79963bd8bc0ae9b13a1aa71add73-Abstract.html>.
- [42] M. Mitzenmacher and S. Vassilvitskii. Algorithms with predictions. In T. Roughgarden, editor, *Beyond the Worst-Case Analysis of Algorithms*, page 646–662. Cambridge University Press, 2021. doi: 10.1017/9781108637435.037.
- [43] M. Mitzenmacher and S. Vassilvitskii. Algorithms with predictions. *Commun. ACM*, 65(7): 33–35, 2022.
- [44] U. G. A. Office. Reverse auctions. <https://www.gao.gov/products/gao-18-446>, 2024. [accessed 3 March 2025].
- [45] M. Purohit, Z. Svitkina, and R. Kumar. Improving online algorithms via ML predictions. In *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*, pages 9684–9693, 2018.

- [46] Y. Shi and Z. Xue. Prediction-augmented mechanism design for weighted facility location. *arXiv preprint arXiv:2507.06509*, 2025.
- [47] D. B. Shmoys, E. Tardos, and K. Aardal. Approximation algorithms for facility location problems (extended abstract). In *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing, STOC '97*, page 265–274, New York, NY, USA, 1997. Association for Computing Machinery. ISBN 0897918886. doi: 10.1145/258533.258600. URL <https://doi.org/10.1145/258533.258600>.
- [48] K. Talwar. The price of truth: Frugality in truthful mechanisms. In *Proceedings of the 20th International Symposium on Theoretical Aspects of Computer Science (STACS)*, pages 608–619. Springer, 2003.
- [49] C. Xu and P. Lu. Mechanism design with predictions. In L. D. Raedt, editor, *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI 2022, Vienna, Austria, 23-29 July 2022*, pages 571–577. ijcai.org, 2022.

A Extended Related Work

Algorithms with Predictions. In recent years, the learning-augmented framework has become a leading paradigm for algorithm design and analysis. For an overview of its early developments, see [43], and for a comprehensive, up-to-date survey, consult [37]. By incorporating (potentially imperfect) predictions, this framework overcomes the limitations of classical worst-case analyses. Indeed, over the past five years, hundreds of papers have revisited core algorithmic problems through this lens, with prominent examples including online paging [39], scheduling [45], covering and knapsack-constrained optimization [10, 30], Nash social welfare maximization [11], variations of the secretary problem [2, 23, 26], and a variety of graph-based challenges [5].

UFL. The metric Uncapacitated Facility Location (UFL) problem is a fundamental NP-hard optimization problem, widely studied due to its theoretical significance and practical relevance in operations research and algorithmic game theory. Early foundational results established the APX-hardness of the problem, showing that it cannot be approximated within a factor better than 1.463 unless $P = NP$ [29]. The first constant-factor approximation was provided by Shmoys et al. [47], who achieved a ratio of approximately 3.16. Subsequent improvements significantly narrowed the approximation gap, with notable progress by Chudak and Shmoys [21] to 1.736 and Mahdian et al. [40] to 1.52-approximation via sophisticated linear-programming techniques. Byrka and Aardal [14] later improved this bound to a 1.50-approximation using a novel bifactor algorithm. The current state-of-the-art algorithm, due to Li [36], achieves an approximation ratio of 1.488, leaving a small gap relative to the known hardness bound.

B Tightness of the Frugality Bound for VCG

We now show that the frugality ratio of 3 for VCG is tight by exhibiting the following instance.

Theorem B.1. *The frugality ratio of VCG auction is at least $3 - \frac{6}{|L|+1}$, even for tree metrics. Hence, in the worst case (as $|L| \rightarrow \infty$), the frugality ratio approaches 3.*

Proof. Consider the tree metric defined by facilities $L = \{\ell_0, \dots, \ell_k\}$, users $U = \{u_1, \dots, u_k\}$, and edges $E = \{\{\ell, u_i\}\}_{i=1}^k \cup \{\{u_i, \ell_0\}\}_{i=1}^k$ (see Figure B for an illustration). Facility ℓ_0 has an opening cost $o_{\ell_0} = 2$ and the other facilities have an opening cost of 0. The distance of from the central facility ℓ_0 to any other facility ℓ is 2 and $d(\ell, \ell_j) = 4$ for all pairs of distinct $i, j > 0$. Finally, we have one user u_i in the midway of ℓ and ℓ_0 .

For the given instance, the optimal solution $\text{OPT}(U, \mathbf{o}, d)$ is to open all of the peripheral facilities $\{\ell_1, \dots, \ell_k\}$ for a total cost of k . Consequently, the frugal solution $F(U, \mathbf{o}) = \{\ell_0\}$ is to open the central facility for a total cost of $k + 2$.

Now consider the payment made by VCG. Note that for all $i \in \{1, \dots, k\}$, if the opening cost of ℓ is $o_\ell = \infty$ (and the other opening costs do not change), then user u_i must connect either to the central

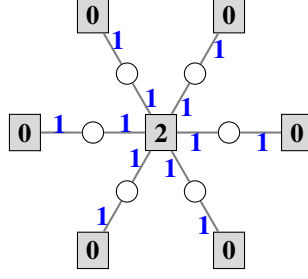


Figure 2: Illustration of the lower bound instance where each square represents a facility location with its opening cost, and the circles represent users.

facility ℓ_0 (which has opening cost 2) or to another peripheral facility (which increases the connection cost of u_i by 2), so the solution $\text{OPT}(\infty, \mathbf{o}_{-l})$ is either $\{\ell_0\}$ or $\{\ell_1, \dots, \ell_{i-1}, \ell_{i+1}, \dots, \ell_k\}$ and $c(\text{OPT}(\infty, \mathbf{o}_{-l})) = k + 2$. Thus, for each peripheral facility ℓ with $i \in \{1, \dots, k\}$, the VCG payment is

$$p_{\ell_i} = c(\text{OPT}(\infty, \mathbf{o}_{-l})) - c(\text{OPT}(0, \mathbf{o}_{-l})) = (k + 2) - k = 2.$$

Therefore, the total payments by the VCG auction to the facilities $\text{OPT}(U, \mathbf{o}) = \{\ell_1, \dots, \ell_k\}$ that it selects is $2k$ which, combined with a total user connection costs of k , leads to a total cost of $3k$. Therefore, the frugality ratio for this instance is

$$\frac{3k}{c(F(U, \mathbf{o}))} = \frac{3k}{k + 2} = \frac{3(|L| - 1)}{|L| + 1} = \frac{3(|L| + 1) - 6}{|L| + 1} = 3 - \frac{6}{|L| + 1}. \quad \square$$

C Missing Proofs from Section 4

C.1 Truthfulness of PREDICTEDLIMITS

We first prove that the auction is truthful. By Myerson's Lemma (Lemma 2.1), it suffices to prove the monotonicity of the allocation rule.

Lemma C.1. *For any $\epsilon \leq 2$, the PREDICTEDLIMITS auction is truthful.*

Proof. We show that the allocation rule of Auction 1 is monotone, and then invoke Myerson's lemma to conclude truthfulness.

Fix a facility $\tilde{\ell} \in L$, and hold constant the true opening costs $c_{-\tilde{\ell}}$ of all other facilities as well as all predictions $\hat{c}$. It suffices to prove that if $\tilde{\ell}$ belongs to the winning set when it reports cost $a \geq 0$, then it still belongs to the winning set when it reports any lower cost $b \leq a$.

Recall that the auction uses the modified cost function

$$c'_\ell(S) = \begin{cases} \frac{2}{\epsilon} c_\ell, & \text{if } S = \text{OPT} \text{ and } c_\ell > \hat{c}_\ell, \\ c_\ell, & \text{otherwise.} \end{cases}$$

We slightly abuse notation and use $o'_\ell(S, a)$ and $o'_\ell(S, b)$ to denote the scaled cost when facility $\tilde{\ell}$ reports a and b , respectively. Since lowering $\tilde{\ell}$'s reported cost can only (weakly) decrease its scaled cost in *every* candidate set, and since the event " $S = \text{OPT}$ " depends only on the prediction, not on the reported bid, we have

$$c'_\ell(S, b) \leq c'_\ell(S, a) \quad \text{for all } S.$$

Let S^* be the winning set when $\tilde{\ell}$ bids a , and let S' be the winning set when it instead bids b , keeping everything else fixed. If $S' = S^*$ then $\tilde{\ell} \in S'$ immediately. Otherwise the only cost that changed is c'_ℓ , so in order for the auction to switch to a different optimal set S' , it must still include $\tilde{\ell}$. Hence in all cases $\tilde{\ell}$ remains selected, proving monotonicity. Finally by Myerson's lemma, a monotone allocation rule induces truthful payments, so the auction is truthful. $\square$

C.2 Missing Proofs from Section 4

C.2.1 Proof of Lemma 4.3

Proof. By definition of the auction, at most one subset, $\hat{\text{OPT}}$, can have its costs scaled. Therefore, it suffices to consider the relationship between $\hat{\text{OPT}}$ and OPT (the true optimal solution). We also note that since $\epsilon \leq 2$, any set that is scaled can only increase its total cost relative to other solutions. The proof structure closely follows the VCG analysis. The main idea is to handle the two cases depending on whether the predicted optimal set $\hat{\text{OPT}}$ matches the true optimal set OPT .

Case one: $\hat{\text{OPT}} \neq \text{OPT}$. In this case, the true optimum OPT was not the set subject to potential scaling. Since $S^* = \text{OPT}$ is the output, OPT must remain the optimal solution even in the modified (potentially scaled) instance. We apply the same VCG analysis structure as in Theorem 3.1 and Lemma 4.2.

Define OPT_f and $\pi_f(\ell)$ as in Theorem 3.1. The bounds from (2) and (3) hold, so we get

$$\begin{aligned}
 \sum_{\ell \in \text{OPT}} \left[p_\ell + \sum_{f: \ell \in \text{OPT}_f} d(U_{\ell,f}, \ell) \right] &\leq \sum_{f \in F} o'_f(F) + 2 \sum_{f \in F} \sum_{\ell \in \text{OPT}_f} d(U_{\ell,f}, f) + \sum_{\ell \in \text{OPT}} \sum_{f: \ell \in \text{OPT}_f} d(U_{\ell,f}, \ell) \\
 &\leq \frac{2}{\epsilon} \sum_{\ell \in F} o_\ell + 2d(U, F) + \sum_{\ell \in \text{OPT}} \sum_{f: \ell \in \text{OPT}_f} d(U_{\ell,f}, \ell) \\
 &\leq \frac{2}{\epsilon} \sum_{\ell \in F} o_\ell + 2d(U, F) + \left(\sum_{\ell \in F} o_\ell + d(U, F) \right) \\
 &= \left(\frac{2}{\epsilon} + 1 \right) \sum_{\ell \in F} o_\ell + 3d(U, F), \tag{5}
 \end{aligned}$$

where the second inequality assumes the worst case for $o'_f(F)$, which occurs if $F = \hat{\text{OPT}}$ and the opening costs of facilities in F were under-predicted. The third inequality uses the fact that the connection cost of the true optimal solution is weakly less than the total cost of the frugal solution.

Case two: $\hat{\text{OPT}} = \text{OPT}$. Since $S^* = \text{OPT}$, the consistency analysis from Lemma 4.2 applies directly, and we have:

$$\sum_{\ell \in \text{OPT}} p_\ell + \sum_{\ell \in \text{OPT}} \sum_{f: \ell \in \text{OPT}_f} d(U_{\ell,f}, \ell) \leq (1 + \epsilon) \cdot c(F) = (1 + \epsilon) \left[\sum_{\ell \in F} o_\ell + d(U, F) \right].$$

The lemma statement follows by taking the maximum of the bounds from the two cases □

C.2.2 Proof of Lemma 4.4

Proof. The proof is very similar to the VCG analysis, replacing the facilities in the optimal solution with the facilities in $S = S^* \setminus F$. In addition, we also need to consider the case where the corresponding frugal facility f is included in S^* , which turns out to be an easier case for bounding rerouting costs.

Consider the case where $S^* \neq \text{OPT}$. First, we note that $\hat{\text{OPT}} = \text{OPT}$, otherwise OPT would be output since scaling a solution only makes it worse. Therefore, both S^* and F are unscaled. Let $S = S^* \setminus F$.

Define U_ℓ , $U_{\ell,f}$, $x_f(\ell)$ and $\pi_f(\ell)$ as in Theorem 3.1, but replacing OPT with S . If $f \in S^*$, then rerouting users from $\ell \in S$ to f is bounded by just connecting the users to f . Therefore, in the worst case, all facilities $f \in F$ are not in S^* . Then, inequalities (2) and (3) hold when replacing OPT with S . Therefore, our total payment is bounded by

$$\begin{aligned}
 \sum_{\ell \in S} \left(p_\ell + \sum_{u \in U_\ell} d(u, \ell) \right) &\leq \left(\sum_{f \in F} o_f + 2 \sum_{f \in F} \sum_{\ell \in S_f} d(U_{\ell,f}, f) \right) + \sum_{\ell \in S} \sum_{f: \ell \in S_f} d(U_{\ell,f}, \ell) \\
 &= \sum_{f \in F} [2d(U_j, f) + o_f] + \sum_{\ell \in S} \sum_{f: \ell \in S_f} d(U_{\ell,f}, \ell).
 \end{aligned}$$

Cancelling the total connection cost term $\sum_{\ell \in S} \sum_{u \in U_\ell} d(u, \ell) = \sum_{\ell \in S} \sum_{f: \ell \in S_f} d(U_{\ell, f}, \ell)$ from both sides:

$$\sum_{\ell \in S} p_\ell \leq \sum_{f \in F} [2d(U_j, f) + o_f] \leq \sum_{\ell \in F} o_\ell + 2d(U, F). \quad \square$$

C.2.3 Proof of Lemma 4.6

Proof. Consider any instance $(U, \mathbf{o}, d)$ and any $\epsilon \leq 2$. Let OPT and F be the optimal solution and frugal solution of the given instance. Let S^* be the output of Auction 1.

Case one: $S^* = \text{OPT}$. If $S^* = \text{OPT}$, by Lemma 4.3 we have:

$$\sum_{\ell \in S^*} p_\ell + d(U, S^*) \leq \left(\frac{2}{\epsilon} + 1\right) \sum_{\ell \in F} o_\ell + 3d(U, F).$$

Case two: $S^* \neq \text{OPT}$. We now consider the case where $S^* \neq \text{OPT}$. First, note that the *connection cost* of S^* is weakly less than or equal to the total cost of any solution K , including the frugal solution F . We have:

$$d(U, S^*) \leq \sum_{\ell \in F} o'_\ell(F) + d(U, F) = \sum_{\ell \in F} o_\ell + d(U, F), \quad (6)$$

where the equality is due to the fact that OPT is the solution that is scaled, by the fact that $S^* \neq \text{OPT}$.

Let $S = S^* \setminus F$ and $S^f = S^* \cap F$. Putting together Lemma 4.4 and Lemma 4.5, we get:

$$\begin{aligned} \sum_{\ell \in S^*} p_\ell + d(U, S^*) &\leq \sum_{\ell \in S} p_\ell + \sum_{\ell \in S^f} p_\ell + d(U, S^*) \\ &\leq \sum_{\ell \in F} o_\ell + 2d(U, F) + \sum_{\ell \in S^f} p_\ell + d(U, S^*) \quad (\text{by Lem. 4.4}) \\ &\leq \max\left(3, \frac{2}{\epsilon} + 1\right) \sum_{\ell \in F} o_\ell + \max\left(4, \frac{2}{\epsilon} + 2\right) d(U, F) + d(U, S^*) \\ &\quad (\text{by Lem. 4.5}) \\ &\leq \max\left(4, \frac{2}{\epsilon} + 2\right) \sum_{\ell \in F} o_\ell + \max\left(5, \frac{2}{\epsilon} + 3\right) d(U, F) \quad (\text{by (6)}) \end{aligned}$$

Taking the worst case of the two scenarios we get that

$$\sum_{\ell \in S^*} p_\ell + d(U, S^*) \leq \max\left(5, \frac{2}{\epsilon} + 3\right) \left[\sum_{i \in F} o_i + d(U, F) \right] \quad \square$$

D Proof of Theorem 4.1

Proof of Theorem 4.1. Auction PREDICTEDLIMITS is truthful by Lemma C.1, $(1 + \epsilon)$ -consistent by Lemma 4.2, and $\max\left(5, \frac{2}{\epsilon} + 3\right)$ -robust by Lemma 4.6. $\square$

E Missing Proofs from Section 5

E.1 Truthfulness of ERRORTOLERANT

We first prove the truthfulness of Auction ERRORTOLERANT. The proof is very similar to Theorem C.1.

Lemma E.1. *For any $\epsilon \leq 2$, the PREDICTEDLIMITS auction is Truthful.*

Proof. We show that the allocation rule of Auction 2 is monotone.

Fix a facility $\tilde{\ell} \in L$, and $\mathbf{o}_{-\tilde{\ell}}$ of all other facilities as well as all predictions $\hat{\mathbf{o}}$. Let $\tilde{\ell}$'s reported cost be $a \geq 0$, and suppose that when it bids a the auction selects $S_a^* \ni \tilde{\ell}$. We must show that if it instead reports any lower cost $b \leq a$, it remains selected.

Denote by $c_\lambda(S; x)$ the total cost of set S under report $x \in \{a, b\}$, according to Auction 2. Note that in either branch (the “whole-solution downscaling” or the “per-facility inflation”), lowering $\tilde{\ell}$'s cost from a to b *weakly decreases* its contribution to $c_\lambda(S)$ for *every* S . We consider two cases:

Case 1: Whole-solution downscaling applies at report a . That is,

$$\forall \ell' \in \hat{\text{OPT}}, o_{\ell'} \leq \lambda \hat{o}_{\ell'} \quad \text{under } x = a.$$

Since $b \leq a$, the same test holds at $x = b$, so the auction remains in the downscaling branch. In that branch

$$c_\lambda(S; x) = \begin{cases} \frac{1}{\lambda^2} \left(d(U, S) + \sum_{\ell \in S} o_\ell \right), & S = \hat{\text{OPT}}, \\ d(U, S) + \sum_{\ell \in S} o_\ell, & \text{otherwise,} \end{cases}$$

and reducing $\tilde{\ell}$'s cost from a to b multiplies its term by at most $b/a \leq 1$ in every candidate set. Hence

$$c_\lambda(S; b) \leq c_\lambda(S; a) \quad \text{for all } S,$$

Since solution that got a reduced cost must contain $\tilde{\ell}$, any new minimizer S_b^* must still contain $\tilde{\ell}$.

Case 2: Per-facility inflation applies at report a . Then there is some $\ell' \in \hat{\text{OPT}}$ with $o_{\ell'} > \lambda \hat{o}_{\ell'}$ under $x = a$. Two subcases arise:

- If $\tilde{\ell} \notin \hat{\text{OPT}}$. Lowering its bid does not affect the branch test, so we stay in the inflation branch. But in that branch each facility's scaled cost (whether inflated by $2/\epsilon$ or not) is a nondecreasing function of its report, so the argument from the original auction applies verbatim to show monotonicity.

- If $\tilde{\ell} \in \hat{\text{OPT}}$. Then $a > \lambda \hat{o}_{\tilde{\ell}}$ but $b \leq \lambda \hat{o}_{\tilde{\ell}}$, so at $x = b$ the auction flips into the downscaling branch. In that branch it will choose $\hat{\text{OPT}}$ (since we uniformly downscale its cost), and $\hat{\text{OPT}} \ni \tilde{\ell}$. Thus $\tilde{\ell}$ remains selected.

In all cases, lowering $\tilde{\ell}$'s bid can only weakly decrease its cost in every candidate set including it, and any switch of the minimizer continues to include $\tilde{\ell}$. Therefore the allocation rule is monotone, and by Myerson's lemma the resulting payments make the auction truthful. $\square$

E.2 Performance analysis for ERRORTOLERANT

We now provide the performance as a function of the error η and the error tolerance parameter λ .

E.2.1 Performance when $\eta \leq \lambda$

We first consider the case where $\eta \leq \lambda$, we start by showing the claim that $\hat{\text{OPT}}$ is always outputted when $\eta \leq \lambda$. The result of this case is summarized in Lemma E.8. We first show that under the condition $\eta \leq \lambda$, the prediction optimal is always returned.

Lemma E.2. *If $\eta \leq \lambda$, then Auction 2 always selects $\hat{\text{OPT}}$; that is, $S^* = \hat{\text{OPT}}$.*

We now provide an upper bound of the connection cost of the output solution.

Proof. Let

$$\hat{c}(S) = d(U, S) + \sum_{\ell \in S} \hat{o}_\ell,$$

be the predicted total cost. First note that since $\eta \leq \lambda$, we have $o_\ell \leq \lambda \hat{o}_\ell$ for all ℓ the auction will set

$$c_\lambda(S) = \begin{cases} \frac{1}{\lambda^2} c(\hat{\text{OPT}}), & S = \hat{\text{OPT}}, \\ c(S), & S \neq \hat{\text{OPT}}. \end{cases}$$

Since $\hat{\text{OPT}}$ minimizes the *predicted* cost,

$$\hat{c}(\hat{\text{OPT}}) \leq \hat{c}(S) \quad \forall S \subseteq L. \tag{7}$$

By $\eta \leq \lambda$ we have $o_\ell \leq \lambda \hat{o}_\ell$ for all $\ell \in \text{OPT}$, so

$$c(\text{OPT}) = \sum_u d(u, \text{OPT}) + \sum_{\ell \in \text{OPT}} o_\ell \leq \sum_u d(u, \text{OPT}) + \lambda \sum_{\ell \in \text{OPT}} \hat{o}_\ell \leq \lambda \hat{c}(\text{OPT}),$$

On the other hand, since $\hat{o}_\ell \leq \eta o_\ell$ for all ℓ , we get

$$\hat{c}(S) = \sum_u d(u, S) + \sum_{\ell \in S} \hat{o}_\ell \leq \sum_u d(u, S) + \eta \sum_{\ell \in S} o_\ell = \eta c(S),$$

Combining the above inequality with (7) we have,

$$c(\text{OPT}) \leq \lambda \hat{c}(S) \leq \lambda \eta c(S) \implies \frac{1}{\lambda^2} c(\text{OPT}) \leq c(S) \implies c_\lambda(\text{OPT}) \leq c_\lambda(S), \quad \forall S \neq \text{OPT}.$$

Since the auction output the set that minimizes c_λ , we therefore have $S^* = \text{OPT}$. $\square$

Observation E.3. For any set $S \neq \text{OPT}$, let $\hat{c}(S) = d(U, S) + \sum_{\ell \in S} \hat{o}_\ell$, we have

$$\hat{c}(S) \leq \eta c(S).$$

Proof. $\eta \leq \lambda$ implies $\hat{o}_\ell \leq \lambda o_\ell$ for every ℓ , so

$$\hat{c}(S) = d(U, S) + \sum_{\ell \in S} \hat{o}_\ell \leq d(U, S) + \eta \sum_{\ell \in S} o_\ell = \eta \left(d(U, S) + \sum_{\ell \in S} o_\ell \right). \quad \square$$

We now provide an upper bound of the connection cost of the output solution.

Lemma E.4. If $\eta \leq \lambda$, then for any $S \subseteq L$,

$$d(U, \text{OPT}) \leq \eta \left(d(U, S) + \sum_{\ell \in S} o_\ell \right).$$

Proof. Since $d(U, \text{OPT}) \leq \hat{c}(\text{OPT})$, combining with the fact that OPT minimizes $\hat{c}$, we have gives

$$d(U, \text{OPT}) \leq \hat{c}(\text{OPT}) \leq \hat{c}(S) \leq \eta \left(d(U, S) + \sum_{\ell \in S} o_\ell \right),$$

where the last inequality is by Observation E.3. $\square$

We now move on to upper bound the payment to the facilities in the output solution, i.e., OPT . In particular, we will partition OPT based on the threshold they applied, and whether it is a member of F , the true frugal solution. We first analyze the low-threshold facilities.

Lemma E.5. Under the condition $\eta \leq \lambda$, let $S^1 = \{\ell \in \text{OPT} : p_\ell \leq \lambda \hat{o}_\ell\}$. Then

$$\sum_{\ell \in S^1} p_\ell \leq \lambda \eta c(F),$$

where F is the frugal solution.

Proof. Since $p_\ell \leq \lambda \hat{o}_\ell$ for every $\ell \in S^1$,

$$\sum_{\ell \in S^1} p_\ell \leq \lambda \sum_{\ell \in S^1} \hat{o}_\ell \leq \lambda \sum_{\ell \in \text{OPT}} \hat{o}_\ell \leq \lambda \hat{c}(\text{OPT}) \leq \lambda \hat{c}(F) \leq \lambda \eta c(F),$$

where the last inequality is by Observation E.3. $\square$

Lemma E.6. Assume $\eta \leq \lambda$, and let

$$S^2 = \{\ell \in \text{OPT} : p_\ell > \lambda \hat{o}_\ell, \ell \notin F\}.$$

Then

$$\sum_{\ell \in S^2} p_\ell \leq \epsilon c(F),$$

where F is the frugal solution.

Proof. We follow the rerouting argument of Lemma 4.4, with the additional fact that when the reported bid is more than $\lambda \hat{o}_\ell$, the opening cost is scaled up. specialized to the output $\hat{\text{OPT}}$.

Define $U_{\ell,f}$ and $\pi_f(\ell)$ as in Theorem 3.1, but replacing OPT with S^2 . If $f \in S^*$, then the cost of rerouting users from $\ell \in S^2$ to f is bounded by just connecting the users to f . Therefore, in the worst case, all facilities $f \in F$ are not in S^* . Then, inequalities (2) and (3) hold when replacing OPT with S . For $\ell \in S^2$, payments are scaled by $\frac{2}{\epsilon}$. Following a similar argument as Lemma 4.4, our total payment is bounded by

$$\begin{aligned} \frac{2}{\epsilon} \sum_{\ell \in S^2} p_\ell + \sum_{\ell \in S^2} \sum_{f: \ell \in S_f^2} d(U_{\ell,f}, \ell) &\leq \sum_{f \in F} \left[2 \sum_{\ell \in S_f^2} d(U_{\ell,f}, f) + o_f + \sum_{\ell \in S_f^2} d(U_{\ell,f}, \ell) \right] \\ &= \sum_{f \in F} [2 d(U_f, f) + o_f] + \sum_{\ell \in S^2} \sum_{f: \ell \in S_f^2} d(U_{\ell,f}, \ell). \end{aligned}$$

Cancelling the total connection cost term $\sum_{\ell \in S^2} \sum_{f: \ell \in S_f^2} d(U_{\ell,f}, \ell)$ from both sides gives

$$\frac{2}{\epsilon} \sum_{\ell \in S^2} p_\ell \leq \sum_{f \in F} [2 d(U_f, f) + o_f],$$

we therefore get

$$\sum_{\ell \in S^2} p_\ell \leq \frac{\epsilon}{2} \sum_{f \in F} o_f + \epsilon \sum_{f \in F} d(U_f, f) \leq \epsilon \cdot c(F). \quad \square$$

Lemma E.7. Under the condition $\eta \leq \lambda$, let

$$S^3 = \{\ell \in \hat{\text{OPT}} : p_\ell > \lambda \hat{o}_\ell, \ell \in F\}.$$

Then

$$\sum_{\ell \in S^3} p_\ell \leq \epsilon \cdot c(\hat{\text{OPT}}).$$

Proof. For $\ell \in S^3$, $\ell \in F$ and $p_\ell > \lambda \hat{o}_\ell$. An argument analogous to Lemma 4.5 shows that inflating the opening cost of any facility in the frugal solution beyond its predicted cost by up to a $2/\epsilon$ factor bounds its threshold payment by ϵ times the cost of the true optimum. Hence $\sum_{\ell \in S^3} p_\ell \leq \epsilon c(\text{OPT})$. $\square$

Lemma E.8. Assume $\eta \leq \lambda$. Let F be the frugal solution, and let S^* and $\{p_\ell\}_{\ell \in S^*}$ be the output of Auction 2. Then

$$d(U, S^*) + \sum_{\ell \in S^*} p_\ell \leq (\eta(1 + \lambda) + 2\epsilon) \cdot c(F).$$

Proof. By Lemma E.2, under $\eta \leq \lambda$ the auction selects $S^* = \hat{\text{OPT}}$. We partition the payments over $\hat{\text{OPT}}$ into three parts S^1, S^2, S^3 as in Lemmas E.5, Lemma E.6 and Lemma E.7. since $c(\text{OPT}) \leq c(F)$. Summing connection cost and all payments gives

$$\begin{aligned} d(U, S^*) + \sum_{\ell \in S^*} p_\ell &= d(U, \hat{\text{OPT}}) + \sum_{\ell \in S^1} p_\ell + \sum_{\ell \in S^2} p_\ell + \sum_{\ell \in S^3} p_\ell \\ &\leq \eta c(F) + \sum_{\ell \in S^1} p_\ell + \sum_{\ell \in S^2} p_\ell + \sum_{\ell \in S^3} p_\ell && \text{(by Lem. E.4)} \\ &\leq \eta c(F) + \lambda \eta c(F) + \sum_{\ell \in S^2} p_\ell + \sum_{\ell \in S^3} p_\ell && \text{(by Lem. E.5)} \\ &\leq \eta c(F) + \lambda \eta c(F) + \epsilon c(F) + \sum_{\ell \in S^3} p_\ell && \text{(by Lem. E.6)} \\ &\leq \eta c(F) + \lambda \eta c(F) + \epsilon c(F) + \epsilon c(F) && \text{(by Lem. E.7)} \\ &= (\eta(1 + \lambda) + 2\epsilon) \cdot c(F). \quad \square \end{aligned}$$

E.2.2 Performance for general error η

We now analyze performance for arbitrary η by breaking the proof into two dimensions: (a) which scaling rule the auction applies: either whole-solution downscaling or per-facility inflation—and (b) which set is ultimately selected. In each of the resulting subcases, we separately bound the connection cost and the total payment using rerouting arguments and the appropriate scaling factor; combining these four analyses yields the unified frugality guarantee stated in Lemma E.11. We start by consider the case where *whole-solution scaling* is applied and the OPT is outputted.

Lemma E.9. *Let S^* be the output of Auction 2, if $S^* = \text{OPT}$, for any $S \neq \text{OPT}$,*

$$d(U, S^*) \leq \lambda^2 c(S).$$

Proof. By the scaling scheme of the auction we have $d(U, \text{OPT}) \leq \lambda^2 c_\lambda(\text{OPT})$, combining with the fact that S^* minimizes c_λ , we have gives

$$d(U, S^*) \leq \lambda^2 \cdot c_\lambda(S^*) \leq \lambda^2 \cdot c_\lambda(S) \leq \lambda^2 \cdot c(S)$$

where the last inequality is since $S \neq \text{OPT}$. $\square$

Lemma E.10. *Let S^* be the output of Auction 2. If $S^* = \text{OPT}$ and the whole-solution-scaling is applied, then*

$$\sum_{\ell \in S^*} p_\ell \leq (2\lambda^4 + 2\lambda^2) c(F),$$

Proof. We follow the rerouting argument of Lemma 4.4. Partition OPT into

$$S = S^* \setminus F \quad \text{and} \quad S^f = \text{OPT} \cap F.$$

Define $U_{\ell,f}$ and $\pi_f(\ell)$ as in Theorem 3.1 but replace OPT with S^*

Since the auction applies a $1/\lambda^2$ down-scaling to OPT , no profitable deviation implies

$$\frac{1}{\lambda^2} \left(p_i + \sum_{j: \ell \in S_f} d(U_{\ell,f}, \ell) \right) \leq \sum_{j: \ell \in S_f} r_{\ell,f}, \quad (8)$$

for each $\ell \in S$, where $r_{\ell,f}$ is the cost to reroute $U_{\ell,f}$ to $\pi_f(\ell)$.

By the standard charging argument,

$$\sum_{\ell \in S_f} r_{\ell,f} \leq 2 \sum_{\ell \in S_f} d(U_{\ell,f}, f) + o_f + \sum_{\ell \in S_f} d(U_{\ell,f}, \ell). \quad (9)$$

Summing (8) over all $\ell \in S$ and then using (9) together with Lemma E.9 ($\sum_f d(U_f, f) \leq d(U, F)$) yields

$$\frac{1}{\lambda^2} \sum_{\ell \in S} p_i \leq 2c(F) + \left(1 - \frac{1}{\lambda^2}\right) \sum_{\ell \in S} \sum_{j: \ell \in S_f} d(U_{\ell,f}, \ell) \leq 2c(F) + (\lambda^2 - 1)c(F).$$

Hence

$$\sum_{\ell \in S} p_\ell \leq (\lambda^4 + \lambda^2) c(F).$$

The case $f \in S^*$ gives the even better bound $r_{\ell,f} \leq d(U_{\ell,f}, f)$, so it cannot increase the total.

Applying the identical rerouting argument to S^f (with F replaced by OPT) yields the same upper bound $(\lambda^4 + \lambda^2)c(F)$. Summing over both parts,

$$\sum_{\ell \in S^*} p_\ell = \sum_{\ell \in S} p_\ell + \sum_{\ell \in S^f} p_\ell \leq (2\lambda^4 + 2\lambda^2) c(F). \quad \square$$

Lemma E.11. *The Auction 2 achieves a frugality ratio*

$$\max\left\{2\lambda^4 + 3\lambda^2, 3 + \frac{2}{\epsilon}\right\}.$$

Proof. The auction uses exactly one of two branches:

(i) **Whole-solution downscaling:**

$$C_\lambda(S) = \begin{cases} \frac{1}{\lambda^2} \left(d(U, S) + \sum_{\ell \in S} o_\ell \right), & S = \hat{\text{OPT}}, \\ d(U, S) + \sum_{\ell \in S} o_\ell, & S \neq \hat{\text{OPT}}. \end{cases}$$

(ii) **Per-facility inflation:**

$$C_\lambda(S) = d(U, S) + \sum_{\ell \in S} o'_\ell(S),$$

where $o'_\ell(S) = \frac{2}{\epsilon} o_\ell$ if $S = \hat{\text{OPT}}$ and $o_\ell > \lambda \hat{o}_\ell$, and $o'_\ell(S) = o_\ell$ otherwise.

Case 1: Whole-solution downscaling applies.

Case 1.1: If the auction still outputs $S^* = \hat{\text{OPT}}$, then by Lemma E.9,

$$d(U, S^*) \leq \lambda^2 c(F).$$

Combined with Lemma E.10, which gives $\sum_{\ell \in S^*} p_\ell \leq (2\lambda^4 + 2\lambda^2) c(F)$, we obtain

$$d(U, S^*) + \sum_{\ell \in S^*} p_\ell \leq (2\lambda^4 + 3\lambda^2) c(F).$$

Case 1.2: Otherwise $S^* \neq \hat{\text{OPT}}$. In this subcase, $\hat{\text{OPT}}$ was downscaled but not chosen, and all other candidate sets remain unscaled. Thus from the winners' perspective the outcome is no worse than standard VCG, so

$$d(U, S^*) + \sum_{\ell \in S^*} p_\ell \leq 3 c(F).$$

Case 2: Per-facility inflation applies. In this branch the analysis of Section 4.3 carries over unchanged: the exact triggering condition ($o_\ell > \hat{o}_\ell$ vs. $o_\ell > \lambda \hat{o}_\ell$) does not affect the worst-case bound. By the robustness proof of Theorem 4.1,

$$d(U, S^*) + \sum_{\ell \in S^*} p_\ell \leq \max\left\{5, 3 + \frac{2}{\epsilon}\right\} c(F).$$

Taking the maximum over all branches and subcases yields

$$\max\left\{2\lambda^4 + 3\lambda^2, 5, 3 + \frac{2}{\epsilon}\right\} = \max\left\{2\lambda^4 + 3\lambda^2, 3 + \frac{2}{\epsilon}\right\} c(F). \quad \square$$

Putting the findings together gives up the performance of `ERRORTOLERANT`.

Proof of Theorem 5.1. If $\eta \leq \lambda$, by Lemma E.8 have the first statement and Lemma E.11 proves the other bound, for arbitrary η values. $\square$

F Computational Complexity

The standard IP formulation for UFL uses binary variables $x_\ell \in \{0, 1\}$ to denote whether facility $\ell \in L$ is opened and variables $y_{u\ell} \in [0, 1]$ for assigning user $u \in U$ to facility ℓ .

The implementation of `Auction 1` proceeds as follows:

1. **Determine the Predicted Optimal Set:** First, we solve a classic UFL integer program using the *predicted* opening costs, $\{\hat{o}_\ell\}_{\ell \in L}$, to find the predicted optimal set, $\hat{\text{OPT}}$.
2. **Find the Best Alternative Solution:** After identifying $\hat{\text{OPT}}$, its modified cost according to the auction rules is calculated directly. To find the best alternative, we solve a second UFL

integer program using the *true* opening costs, $\{o_\ell\}_{\ell \in L}$. A linear constraint is added to this IP to make the set $\hat{OPT}$ an infeasible solution:

$$\sum_{\ell \in \hat{OPT}} x_\ell \leq |\hat{OPT}| - 1$$

This inequality ensures that at least one facility from the set $\hat{OPT}$ remains unopened, allowing the IP to find the optimal solution among all other feasible sets.

3. **Select the Winning Set:** The auction's final output, S^* , is determined by comparing the modified cost of $\hat{OPT}$ with the cost of the best alternative solution found in the previous step. The set with the lower cost is chosen.
4. **Compute Payments:** Finally, for each facility ℓ in the winning set S^* , its threshold payment is computed. This is done by solving one additional UFL integer program per winning facility to find the critical cost at which that facility would no longer be part of the optimal solution.

NeurIPS Paper Checklist

1. Claims

Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

Answer: [\[Yes\]](#)

Justification: Our primary claims are establishing a tight bound for the VCG mechanism frugality ratio in the incapacitated facility location problem, and introduce a prediction-enhanced mechanism that achieves good consistency and robustness guarantees. Both of these claims are achieved in the paper and proven in Sections 3 and 4 respectively.

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. Limitations

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [\[Yes\]](#)

Justification: we have clearly stated our mathematical model and the assumptions we make about the model, including the metric space, types of predictions and explicitly explained that the results hold under the given model.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. Theory assumptions and proofs

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: The main goal of the paper is to prove claims about the frugality ratio of several mechanisms for the incapacitated facility location. Therefore, all theorems and lemmas provided in the paper are stated with appropriate assumptions and justified with a complete proof.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. Experimental result reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [NA]

Justification: The paper does not include experiments. All results are entirely theoretical.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
 - (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
 - (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
 - (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
 - (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in

some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [NA]

Justification: The paper does not include any experiments, nor does it use on data or code.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. Experimental setting/details

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [NA]

Justification: The paper does not include experiments.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. Experiment statistical significance

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [NA]

Justification: The paper does not include experiments.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. Experiments compute resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [NA]

Justification: The paper does not include experiments.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. Code of ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines>?

Answer: [Yes]

Justification: The authors have reviewed the Code of Ethics and concluded that the research conforms with all guidelines listed.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. Broader impacts

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [NA]

Justification: The paper presents purely theoretical results (e.g., mathematical theorems and performance bounds) without proposing deployable systems, datasets, or algorithms. Consequently, we do not anticipate any direct positive or negative societal impacts that warrant further discussion.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. Safeguards

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: The paper does not use models or data.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. Licenses for existing assets

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [NA]

Justification: The paper does not use existing assets.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.

- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [NA]

Justification: The paper does not release new assets.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.

- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: The paper does not use LLMs.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (<https://neurips.cc/Conferences/2025/LLM>) for what should or should not be described.