

# Set Covering with Our Eyes Wide Shut\*

Anupam Gupta<sup>†</sup>

Gregory Kehne<sup>‡</sup>

Roie Levin<sup>§</sup>

## Abstract

In the *stochastic set cover problem* (Grandoni et al., FOCS '08), we are given a collection  $\mathcal{S}$  of  $m$  sets over a universe  $\mathcal{U}$  of size  $N$ , and a distribution  $D$  over elements of  $\mathcal{U}$ . The algorithm draws  $n$  elements one-by-one from  $D$  and must buy a set to cover each element on arrival; the goal is to minimize the total cost of sets bought during this process. A *universal* algorithm a priori maps each element  $u \in \mathcal{U}$  to a set  $S(u)$  such that if  $U \subseteq \mathcal{U}$  is formed by drawing  $n$  times from distribution  $D$ , then the algorithm commits to outputting  $S(U)$ . Grandoni et al. gave an  $O(\log mN)$ -competitive universal algorithm for this stochastic set cover problem.

We improve unilaterally upon this result by giving a simple, polynomial time  $O(\log mn)$ -competitive universal algorithm for the more general *prophet* version, in which  $U$  is formed by drawing from  $n$  different distributions  $D_1, \dots, D_n$ . Furthermore, we show that we do not need full foreknowledge of the distributions: in fact, a single sample from each distribution suffices. We show similar results for the *2-stage prophet* setting and for the *online-with-a-sample* setting.

We obtain our results via a generic reduction from the single-sample prophet setting to the random-order setting (for which Gupta et al., FOCS 2021 provides an algorithm); this reduction holds for a broad class of minimization problems that includes all covering problems. We take advantage of this framework by giving random-order algorithms for non-metric facility location and set multicover; using our framework, these automatically translate to universal prophet algorithms.

## 1 Introduction

In the SETCOVER problem we are given a set system  $(\mathcal{U}, \mathcal{S})$ , where  $\mathcal{U}$  is a ground set of size  $N$  and  $\mathcal{S}$  is a collection of subsets with  $|\mathcal{S}| = m$ . We are also given a subset  $U \subseteq \mathcal{U}$  of size  $n$ . The goal is to select a minimum-size (or more generally, minimum-cost) subcollection  $\mathcal{S}' \subseteq \mathcal{S}$  such that the union of the sets in  $\mathcal{S}'$  is  $U$ . Many polynomial-time algorithms have been discovered for this problem that achieve an approximation ratio of  $\ln n$  (see e.g. [Chv79, Joh74, Lov75, WS11]), and this is best possible unless  $P = NP$  [Fei98, DS14].

One may interpret a solution  $\mathcal{S}'$  as a map  $\mathfrak{S} : \mathcal{U} \rightarrow \mathcal{S}$  taking each element to a set that covers it (breaking ties arbitrarily). In this case  $\mathfrak{S}(U) = \bigcup_{u \in U} \{\mathfrak{S}(u)\}$  is the solution  $\mathcal{S}'$ . In seminal work, Jia et al. defined the *universal* variant of the set cover problem, in which the goal is to construct  $\mathfrak{S}$  a priori and obliviously without seeing the actual value of  $U \subseteq \mathcal{U}$  (hence it is constructed using only  $\mathcal{U}$  and  $\mathcal{S}$ ) [JLN+05]. One wants a map  $\mathfrak{S}$  minimizing the worst case ratio  $\max_{U \subseteq \mathcal{U}} c(\mathfrak{S}(U))/c(\text{OPT}(U))$  between the cost of  $\mathfrak{S}(U)$  and the cost of the optimal set cover for  $U$ . A universal algorithm is said to be  $\alpha$ -competitive, or to achieve competitive ratio  $\alpha$ , if the value of this ratio is no more than  $\alpha$ . Jia et al. showed  $\tilde{O}(\sqrt{n})$  bounds for this problem [JLN+05].

To overcome this polynomial barrier, Grandoni et al. [GGL+13] studied the *stochastic* variant of universal set cover, in which one additionally assumes that the elements of  $U$  are drawn i.i.d. from a known distribution  $D$ . The aim is now to minimize the expected ratio  $E_U[c(\mathfrak{S}(U))]/E_U[c(\text{OPT}(U))]$ . With this assumption, they showed that it is possible to get an exponentially better  $O(\log(mN))$ -competitive algorithm, and that this is best possible up to log log factors.

In this work, we improve, generalize and simplify the results of [GGL+13]. First, we improve the competitive ratio to  $O(\log(mn))$ , which can be exponentially smaller when  $n \ll N$ . This essentially is best possible for polynomial-time algorithms, since there is an  $\Omega(\log m / \log \log m)$  lower bound when  $n \ll m$  [GGL+13], and there is no polynomial-time algorithm with approximation  $o(\log n)$  unless  $P = NP$  [Fei98, DS14]. We also generalize to the *prophet* setting, in which  $U$  consists of draws from a sequence of non-identical distributions  $D^1, \dots, D^n$ .

\*The full version of the paper can be accessed at <https://arxiv.org/abs/2304.02063>.

<sup>†</sup>Carnegie Mellon University and Google Research. Email: [anupamg@cs.cmu.edu](mailto:anupamg@cs.cmu.edu). Research partly supported by NSF awards CCF-1955785, CCF-2006953, and CCF-2224718.

<sup>‡</sup>Department of Computer Science, University of Texas at Austin, Austin, TX. Email: [gkehne@utexas.edu](mailto:gkehne@utexas.edu).

<sup>§</sup>Department of Statistics and Operations Research, School of Mathematical Sciences, Tel Aviv University, Tel Aviv. Email: [roi1@tauex.tau.ac.il](mailto:roi1@tauex.tau.ac.il).

In fact, we show that we do not need full knowledge of these distributions, and even a single sample from each distribution suffices. We also show extensions to two other related models, the *2-stage* prophet setting, and the online-with-a-sample setting, as well as to several problems which generalize SETCOVER and covering. We now present a more formal overview of these results.

**1.1 Our Results** Our main contribution is a reduction from the prophet setting to the random-order online setting. In random-order SETCOVER, the elements  $U$  are adversarially chosen and revealed one at a time. The algorithm must choose a set to cover the element, and decisions are irrevocable. Since the LEARNORCOVER algorithm of [GKL21] is an  $O(\log(mn))$ -competitive algorithm for random-order SETCOVER, we immediately obtain:

**THEOREM 1.1.** *There is a polynomial-time  $O(\log(mn))$ -competitive universal algorithm for 1-sample prophet SETCOVER.*

Using similar techniques, we obtain theorems for the following two models as well. In the 2-stage prophet model, the algorithm is allowed to purchase sets at a discount in a first stage before the game begins. In a second stage,  $U$  is drawn at random as in the usual prophet setting, and any sets bought after seeing the realizations cost full price.

**THEOREM 1.2.** *There is a polynomial-time universal algorithm for 2-stage prophet SETCOVER that is  $O(\log(mn))$ -competitive with respect to the optimal online policy.*

In the *online-with-a-sample* setting, an adversary selects an unknown element set  $U$  and reveals a uniformly random  $\alpha$ -fraction of it to the algorithm. After this point, the remaining elements are revealed one-by-one in adversarial order. The algorithm must buy sets to cover incoming elements immediately on arrival, and decisions are irrevocable.

**THEOREM 1.3.** *For every  $0 < \alpha \leq 1$ , there is a polynomial-time  $O(\log(mn)/\alpha)$ -competitive universal algorithm for online-with-a-sample SETCOVER.*

In fact our reduction holds for a more general class of minimization problems which we call *augmentable integer programs* (AIPs). These are problems for which augmenting any infeasible solution  $z$  to a feasible solution gets cheaper as  $z$  grows<sup>1</sup>. Hence we reduce the task of designing prophet/2-stage prophet/online-with-a-sample algorithms for AIPs to the task of designing random-order algorithms. To illustrate our reductions, we additionally give  $O(\log mn)$ -competitive algorithms for NONMETRICFACILITYLOCATION and SETMULTICOVER in random order. This marks partial progress in answering an open question of [GKL21], which asked if there is an  $O(\log mn)$ -competitive algorithm for covering integer programs with box constraints in random order.

**1.2 Techniques and Overview** The proofs of Theorems 1.1, 1.2, and 1.3 share a common template. The main idea is to reduce from setting  $\mathcal{X}$  to random-order SETCOVER as follows:

1. Generate a mock instance  $\hat{I} \sim \mathcal{X}$ .
2. Simulate algorithm  $\mathcal{A}$  for random-order SETCOVER on the mock input  $\hat{I}$  by shuffling the order artificially.
3. Solve  $\mathcal{X}$  on the real input  $\mathcal{I} \sim \mathcal{X}$  by first buying the solution  $z$  bought by  $\mathcal{A}$ , then covering any outstanding uncovered element  $v^t$  with the cheapest set containing  $v^t$ .

The idea is to charge the “backup” sets bought to cover any elements missed by  $\mathcal{A}$  to the actual decisions made by  $\mathcal{A}$ , which we can bound using the performance guarantees on  $\mathcal{A}$ . This perspective allows us to give us proofs that are simple in hindsight; yet, prior to our work it was not known how to obtain such results (in fact incorrect proofs of weaker statements were claimed and retracted, as we detail in the related work section and appendix). We treat each model separately in Sections 3, 4, and 5.

<sup>1</sup>Such problems were studied by [GGLS08] but were not given a name. These are intuitively related to *covering* problems; we reserve the term covering for problems with upward closed feasible regions. See Section 2 for details.

In Sections 6 and 7 we give random-order algorithms for `NONMETRICFACILITYLOCATION` and `SETMULTICOVER`. This demonstrates the generality of our reductions, and also illustrates the versatility of the “Learn or Cover” framework of [GKL21] beyond pure covering problems. These are the most technically involved sections of this work.

These results build upon the `LEARNORCOVER` framework of [GKL21] designed for set cover. With every element that arrives uncovered, `LEARNORCOVER` (a) samples from a distribution over sets, and (b) learns from the fact that a random element was uncovered to update the distribution. [GKL21] show the algorithm either makes progress *learning* about the optimal distribution from which one should be sampling, or if it does not then it makes progress *sampling* since the distribution is already sufficiently good. They used a two-part potential, where the parts measure progress learning and covering respectively.

Non-metric facility location is often treated as an extension of set cover, since there are standard reductions between the two ([Vyg05, Section 3.1] and [KT84]). However, these reductions do not hold in the random-order model. The first reduction from `NONMETRICFACILITYLOCATION` to `SETCOVER`, which is folklore, requires an exponential blowup in the number of sets. This is prohibitive since one must in general lose a  $\Omega(\log m)$  factor for random-order online set cover [GKL21]. The second reduction introduces a new set and a new element for every facility-client pair; thus a client arriving in random order becomes a batch of new sets and elements. However, online set cover in which elements arrive in randomly ordered *batches* is in general harder than true uniform random order since [GKL21] show a log-squared lower bound for this problem. Therefore both reductions face obstacles in the random-order setting, and a new approach is needed.

The primary challenge in random-order `FACILITYLOCATION` is to account for *connection costs*. This makes the task of learning a distribution over facilities complex, since the cost of satisfying an arriving client changes over time. Our approach may be viewed as running `LEARNORCOVER` on a set system that evolves dynamically over time: each facility is a set, each client is an element, and a client’s element is contained in a facility’s set if opening that facility significantly reduces that client’s connection cost. We reuse the high level learn/cover idea, but we need to use a more intricate potential to measure progress learning.

Finally, our random-order `SETMULTICOVER` algorithm builds on the slightly more involved algorithm of [GKL21] for random-order CIPs. This involves several technical challenges. For one, the two-part potential of [GKL21] expects that if a variable’s probability in the maintained distribution is high, then it will contribute towards covering unseen constraints in expectation. However multiplicity constraints prohibit the algorithm from sampling any variable more than once, even in this case. We show that this difficulty can nevertheless be circumvented by gradually “forgetting” coordinates that have already hit their caps; interestingly our multiplicative weights update rule does not depend on the marginal augmentation cost of the incoming constraint, as it does in [GKL21].

**1.3 Related Work** The term *prophet inequality* is usually used in the context of online max finding: a gambler draws numbers one-by-one from a sequence of known distributions, and their task is to stop at the highest number. Prophet inequality refers to the bound on the performance of such a player in terms of that of a clairvoyant “prophet” who can see the future. [KS78] showed a strategy for this game with expected reward at least  $1/2$  that of the prophet (see [HK92] for a further survey). The *secretary problem* [F<sup>+</sup>89] is a related max-finding game in which the gambler sees *arbitrary* numbers in *random* order, and once again aims to stop at the highest number. [AKW14] gave a  $1/e$ -competitive strategy for the prophet problem (and extensions) via a reduction to (a subclass of algorithms for) the secretary problem, and this bound was later improved to  $1/2$  by [RWW20]. Our main result may be viewed as a minimization counterpart of the prophet-to-secretary reductions of [AKW14] for maximization problems.

*Free-order* prophet inequalities, in which the gambler can adaptively choose the order in which to open boxes, were studied by [LLP<sup>+</sup>21, PT22, BC22]. Our main result implies that for covering problems, the constrained-order prophet problem is—up to a factor of two—no harder than its free-order counterpart.

Previous work of [DEH<sup>+</sup>18a] claimed a reduction from the prophet set cover problem to universal algorithms for the stochastic (in other words i.i.d.) set cover problem. However the proof (which appears in Section 9.5 of [EB17]) has an issue which we detail in Appendix C, and the claim has since been withdrawn [Sed].

Motivated by settings where an algorithm has access to historical data, [KNR22, KNR20] recently introduced the online-with-a-sample model in the context of max-finding (i.e. the secretary problem), and matching. [AFGS22] study Steiner tree, facility location and load balancing in this model.

There is considerable work on 2-stage (and more generally multi-stage) stochastic optimization from the perspective of approximations (e.g., [SS06a, CCP05, GPRS04, GPRS05, GRS07]), see [SS06b, BL11] for surveys. Our 2-stage prophet model is a *hybrid* stochastic-online model in which the second stage is a fully online game; as far as we know, this model has not been previously studied.

Finally, using our reduction framework and the 3-competitive random-order algorithm of [KNR23], we automatically get 6-competitive universal algorithm for the prophet *metric* facility location problem with a single sample per distribution. A similar result for the special case where all the distributions are identical is implied by previous work of [GGLS08].

## 2 Preliminaries

All logarithms in this paper are taken to be base  $e$ . In the following definitions, let  $x, y \in \mathbb{R}_+^n$  be vectors. The standard dot product between  $x$  and  $y$  is denoted  $\langle x, y \rangle = \sum_{i=1}^n x_i y_i$ . We use  $\max(x, y)$  to denote the coordinate-wise maximum. We use a weighted generalization of KL divergence. Given a weight function  $c$ , define

$$\text{KL}_c(x \parallel y) := \sum_{i=1}^n c_i \left[ x_i \log \left( \frac{x_i}{y_i} \right) - x_i + y_i \right].$$

**Augmentable Integer Programs.** A *covering* integer program (CIP) is usually defined as an integer program (IP) for which the set of feasible solutions is upwards closed. We define the following more general class of problems which we call *augmentable integer programs* (AIPs). These were studied in [GGLS08], but not given an explicit name.

Let  $V$  be a set of requests. For any subset of requests  $V' \subseteq V$ , let  $\text{SOLS}(V') \subseteq \mathbb{Z}^m$  be the subset of solutions that are feasible to  $V'$ . Next, for any subset of requests  $V'$ , any solution  $z \in \text{SOLS}(V')$ , and any request set  $W$ , define the augmentation cost

$$\text{AUG}(W \mid z, V') = \min_w \{c(w) \mid \max(w, z) \in \text{SOLS}(V' \cup W)\},$$

or  $\infty$  if no such  $w$  exists. Let  $\text{BACKUP}(W \mid z, V')$  be a minimizer when it exists.

**DEFINITION 2.1.** (AIP) *An augmentable integer linear program (AIP) is one in which augmentation costs are monotone, i.e. for any  $V' \subseteq V'' \subseteq V$ , and any  $z' \leq z''$  such that  $z' \in \text{SOLS}(V')$  and  $z'' \in \text{SOLS}(V'')$ , we have  $\text{AUG}(W \mid z'', V'') \leq \text{AUG}(W \mid z', V')$  for any request set  $W \subseteq V$ .*

For a set of requests  $V$ , let  $\text{OPT}(V)$  be a minimum-cost solution in  $\text{SOLS}(V)$ , and  $c(\text{OPT}(V))$  its associated cost.

**OBSERVATION 2.1.** (AIPS ARE SUBADDITIVE) *For any  $A, B \subseteq V$ , we have that  $c(\text{OPT}(A \cup B)) \leq c(\text{OPT}(A)) + c(\text{OPT}(B))$ .*

*Proof.* We have that

$$c(\text{OPT}(A \cup B)) \leq c(\text{OPT}(A)) + \text{AUG}(B \mid \text{OPT}(A), A) \leq c(\text{OPT}(A)) + c(\text{OPT}(B)).$$

The first inequality follows since building a solution feasible to  $A$  and then augmenting it to satisfy  $B$  is only more expensive than  $\text{OPT}(A \cup B)$ . The second inequality follows from the monotonicity of augmentation costs property of AIPs, with  $z' = \vec{0}$ ,  $z'' = \text{OPT}(A)$ ,  $V' = \emptyset$ ,  $V'' = A$ , and  $W = B$ .  $\square$

Note that the standard IP formulation of `NONMETRICFACILITYLOCATION` with indicator variables for  $\{x_f\}_f$  for facilities, and  $\{y_{fc}\}_{f,c}$  for facility-client connections is an AIP, but not a CIP. Likewise, `SETMULTICOVER` is an AIP but not a CIP.

**Online Models.** We briefly catalogue the various models that we treat in this paper.

1. An *online* AIP is an AIP in which some constraints are given upfront, and some are revealed sequentially over time. The algorithm must maintain a monotonically increasing solution that satisfies all the constraints revealed so far.

2. A *prophet* AIP instance is an online AIP instance in which the constraints  $v^1, \dots, v^n$  are drawn from known distributions  $D^1, \dots, D^n$ .
3. A *k-sample prophet* AIP instance is an online AIP in which the constraints  $v^1, \dots, v^n$  are drawn from *unknown* distributions  $D^1, \dots, D^n$ , except the algorithm is given  $k$  samples from each of the distributions before the online sequence begins.
4. A *free-order* (resp. *k-sample*) prophet AIP instance is a (resp. *k-sample*) prophet AIP instance in which the algorithm is allowed to adaptively decide the order in which it samples the known (resp. unknown but sampled  $k$  times) distributions  $D^1, \dots, D^n$ .
5. A *2-stage* prophet AIP instance is a prophet AIP instance and a positive number  $\lambda$ . The algorithm is allowed to purchase an initial solution  $z_1$  before the online sequence begins, and any purchases  $z_2$  made during the online sequence are marked up by a factor of  $\lambda$ .
6. An *online-with-a-sample* AIP instance is an online AIP instance in which, after the adversary fixes the input, the algorithm is given a uniformly random  $\alpha$  fraction of the sequence upfront before the online game begins.

### 3 Universal Prophet Algorithms

In this section we prove Theorem 1.1 via a reduction to random-order set cover. Recall that [AKW14] gave such a reduction for maximization problems. Our work offers complimentary results for minimization problems.

In fact, our reduction is more powerful in two ways:

- It holds for AIPs generally, beyond set cover. Hence to construct a prophet algorithm for an AIP, it suffices to construct a random-order algorithm.
- We require a weaker property even than random order. In fact, we can reduce the prophet setting to the *free-order* prophet setting, where the algorithm is granted the freedom to choose the order in which it samples the distributions  $D_1, \dots, D_n$ . Random-order algorithms are a special class of free-order algorithms.

**THEOREM 3.1.** *Let  $\mathcal{I}$  be an instance class of prophet AIP. If algorithm  $\mathcal{A}$  is a free-order prophet AIP algorithm that achieves competitive ratio  $\Delta$  on class  $\mathcal{I}$  using  $k$  samples, then there is a fixed-order prophet AIP algorithm  $\mathcal{A}'$  for class  $\mathcal{I}$  achieving competitive ratio  $2\Delta$  using  $k + 1$  samples.*

In particular, random-order algorithms are 0-sample free-order prophet algorithms, and furthermore LEARNORCOVER of [GKL21] is a random-order set cover algorithm. Hence we get Theorem 1.1 as a corollary.

*Proof of Theorem 3.1.* Let  $\mathcal{A}$  be a  $k$ -sample algorithm for prophet AIP using custom/adaptive order  $\pi$  and with expected competitive ratio  $\Delta$ . Define  $\mathcal{A}'$  to be Algorithm 1.

---

#### Algorithm 1 ( $k + 1$ )-SAMPLE PROPHET TO $k$ -SAMPLE FREE ORDER PROPHET

---

- 1: Train algorithm  $\mathcal{A}$  on  $k$  samples each of  $D^1, \dots, D^n$ .
  - 2: Let  $\text{MOCKRUN} = \{\hat{v}^1, \dots, \hat{v}^n\}$  be one sample each from  $D^1, \dots, D^n$ .
  - 3: **for**  $\tau = 1, 2, \dots, n$  **do**
  - 4:      $\pi^\tau \leftarrow$  the  $\tau^{\text{th}}$  sample in (possibly adaptive) order  $\pi$  specified by  $\mathcal{A}$ .
  - 5:     Feed  $\pi^\tau$  to  $\mathcal{A}$ .
  - 6: Let  $\hat{z}$  be the output of  $\mathcal{A}$ .
  - 7: Initialize  $z \leftarrow \hat{z}$ .
  - 8: **for**  $t = 1, 2, \dots, n$  **do**
  - 9:     Draw  $v^t \sim D^t$ .
  - 10:     **if**  $v^t$  not satisfied by  $z$  **then**
  - 11:         Update  $z \leftarrow \max(z, \text{BACKUP}(v^t \mid z, \text{MOCKRUN} \cup \{v^1, \dots, v^{t-1}\}))$ .
  - 12: **return**  $z$ .
- 

Clearly  $\mathcal{A}'$  uses  $k + 1$  samples, since that is enough samples to simulate  $\mathcal{A}$ : algorithm  $\mathcal{A}$  requires  $k$  samples upfront, and  $\mathcal{A}'$  uses one more to simulate the real draw from each distribution. We turn to bounding the cost of  $\mathcal{A}$ .

The two sets of samples  $v_1, \dots, v_n$  and  $\hat{v}^1, \dots, \hat{v}^n$  are identically distributed, so

$$\mathbb{E}[c(\text{OPT}(v_1, \dots, v_n))] = \mathbb{E}[c(\text{OPT}(\hat{v}^1, \dots, \hat{v}^n))].$$

Thus the expected cost of the solution  $\hat{z}$  bought by  $\mathcal{A}$  is at most  $\Delta \cdot \mathbb{E}[c(\text{OPT}(v_1, \dots, v_n))]$ , by the guarantee on  $\mathcal{A}$ . It remains to bound the cost of the backup purchases in Line 11. To this end, consider each pair of requests  $v^t, \hat{v}^t \sim D^t$ , where  $v^t$  is drawn on Line 9, and  $\hat{v}^t$  is part of the mock run specified on Line 4. We will refer to these pairs of requests as *mates*.

We will argue that the expected augmentation cost of a request  $v^t$  is no more than the expected augmentation cost of its mate  $\hat{v}^t$  during the simulation of  $\mathcal{A}$ . Towards this, let  $z(v)$  be the state of  $z$  at the beginning of the round in which request  $v$  arrives. Let  $\hat{z}(\hat{v})$  be the state of the solution of algorithm  $\mathcal{A}$  at the beginning of round in which  $\hat{v}$  arrives in the simulation of  $\mathcal{A}$  (that is, *in the order  $\pi$  chosen by  $\mathcal{A}$* ). Finally, let  $\text{MOCKRUN}_{<\hat{v}}$  be the set of requests of  $\text{MOCKRUN}$  that arrive before  $\hat{v}$  according to order  $\pi$ . Now:

$$\begin{aligned} \mathbb{E}[\text{AUG}(v^t \mid z(v^t), \text{MOCKRUN} \cup \{v^1, \dots, v^{t-1}\})] &\leq \mathbb{E}[\text{AUG}(v^t \mid \hat{z}(\hat{v}^t), \text{MOCKRUN}_{<\hat{v}^t})] \\ &= \mathbb{E}[\text{AUG}(\hat{v}^t \mid \hat{z}(\hat{v}^t), \text{MOCKRUN}_{<\hat{v}^t})] \end{aligned}$$

The inequality holds by the monotonicity of augmentation property of AIPs in Definition 2.1, since  $z(v^t) \geq \hat{z}(\hat{v}^t)$ . The equality holds because  $v^t$  and  $\hat{v}^t$  are identically distributed. Summing over  $t$ , and noting that  $\mathcal{A}$  must pay at least  $\text{AUG}(\hat{v}^\tau \mid \hat{z}(\hat{v}^\tau), \text{MOCKRUN}_{<\hat{v}^\tau})$  in round  $\tau$  in the event that  $\hat{v}^\tau$  is unsatisfied on arrival, we get that the total backup cost is bounded as

$$\begin{aligned} \sum_t \mathbb{E}[\text{AUG}(v^t \mid z(v^t), \text{MOCKRUN} \cup \{v^1, \dots, v^{t-1}\})] &\leq \sum_t \mathbb{E}[\{\text{AUG}(\hat{v}^t \mid \hat{z}(\hat{v}^t), \text{MOCKRUN}_{<\hat{v}^t})\}] \\ &\leq \mathbb{E}[c(\hat{z})], \end{aligned}$$

which we can bound by  $\Delta \cdot \mathbb{E}[c(\text{OPT}(v_1, \dots, v_n))]$  by assumption of  $\mathcal{A}$ . Thus, in total, Algorithm 1 pays at most  $2 \cdot \Delta \cdot \mathbb{E}[c(\text{OPT})]$ .  $\square$

We note that this proof only requires the monotone augmentation cost property for individual requests.

#### 4 Two-Stage Prophet Algorithms

Recall the 2-stage prophet setting. At the outset we have sample access to distributions  $D_1, \dots, D_t$ , as well as some  $\lambda > 0$ .

- Stage 1: The algorithm may buy a partial solution  $z_0$  and incur cost  $c(z_0)$ .
- Stage 2: Requests  $v_t \sim D_t$  arrive one-at-a-time and the algorithm must augment its solution to satisfy them immediately. If  $z_1$  is the portion of the solution bought in this second phase, the algorithm incurs an additional cost of  $\lambda \cdot c(z_1)$ .

Note that for  $0 < \lambda \leq 1$  the algorithm should always wait to buy sets online, and this reduces to the prophet setting above. For  $\lambda > 1$  we may assume without loss of generality that  $\lambda$  is an integer (at the expense of a small constant factor).

Our aim is to compete with  $\text{OPT}_{\text{ONLINE}}$ , the solution bought by the optimal online algorithm for this two-stage problem. We write  $\text{OPT}_{\text{ONLINE}} = \max(z_0^*, z_1^*)$  where the solution bought in advance  $z_0^*$  is deterministic, and the solution bought during the online sequence  $z_1^*$  depends on the realizations of the draws from the distributions.

Our main result in this section is that the 2-stage setting is no harder than the random-order setting.

**THEOREM 4.1.** *Let  $\mathcal{I}$  be an instance class of prophet AIP. If algorithm  $\mathcal{A}$  is a random-order AIP algorithm that achieves competitive ratio  $\Delta$  on class  $\mathcal{I}$ , then there is a 2-stage prophet AIP algorithm  $\mathcal{A}'$  for class  $\mathcal{I}$  achieving competitive ratio  $2\Delta$  with respect to the optimal online policy using  $\lambda$  samples.*

*Proof.* Let  $\mathcal{A}$  be the random-order algorithm for instance class  $\mathcal{I}$  with expected competitive ratio  $\Delta$ . Define  $\mathcal{A}'$  to be Algorithm 2.

---

**Algorithm 2** 2-STAGE PROPHET TO RANDOM ORDER

---

**First Stage:**

- 1: **for**  $i = 1, \dots, \lambda$  **do**
- 2: | Let  $\text{MOCKRUN}_i \leftarrow \{\widehat{v}_i^1, \dots, \widehat{v}_i^n\}$  be one sample each from  $D^1, \dots, D^n$ .
- 3: Let  $\text{MOCKRUN} = \bigcup_i \text{MOCKRUN}_i$ .
- 4: **for**  $\tau = 1, 2, \dots, \lambda \cdot n$  **do**
- 5: |  $\pi^\tau \leftarrow$  the  $\tau^{\text{th}}$  request of  $\text{MOCKRUN}$  in random order  $\pi$ .
- 6: | Feed  $\pi^\tau$  to  $\mathcal{A}$ .
- 7: Let  $z_0$  be the output of  $\mathcal{A}$ . Buy  $z_0$ .

**Second Stage:**

- 8: Initialize  $z_1 \leftarrow 0$ .
  - 9: **for**  $t = 1, 2, \dots, n$  **do**
  - 10: | Draw  $v^t \sim D^t$ .
  - 11: | Let  $z \leftarrow \max(z_0, z_1)$ .
  - 12: | **if**  $v^t$  not satisfied by  $z$  **then**
  - 13: | | Let  $w \leftarrow \text{BACKUP}(v^t \mid \max(z_0, z_1), \text{MOCKRUN} \cup \{v^1, \dots, v^{t-1}\})$ .
  - 14: | | Update  $z_1 \leftarrow \max(z_1, w)$ .
  - 15: **return**  $\max(z_0, z_1)$ .
- 

The proof proceeds in two steps. Let  $Z := \mathbb{E}[c(z_0^*) + \lambda \cdot c(z_1^*)]$  be the expected cost of  $\text{OPT}_{\text{ONLINE}}$ . First we bound the expected cost of  $z_0$  computed by  $\mathcal{A}$  in terms of  $Z$ , and then we bound the total cost of backups, i.e.  $z_1$ , in terms of the cost of  $z_0$ .

For the first bound on  $z_0$ , we follow the “boosted sampling” argument of [GPRS04]. Suppose that  $z_0^*$  is the optimal first-stage solution, and  $z_1^*, \dots, z_\lambda^*$  are the second-stage solutions bought by the optimal online strategy when fed each of the sequences  $\text{MOCKRUN}_i$ . By the subadditivity property of AIPs from Observation 2.1,

$$c(\text{OPT}(\text{MOCKRUN})) \leq c(z_0^*) + \sum_{i=1}^{\lambda} c(z_i^*) = c(z_0^*) + \mathbb{E}_{i \sim [\lambda]} [\lambda \cdot c(z_i^*)],$$

where  $i \sim [\lambda]$  denotes that  $i$  is drawn uniformly from  $[\lambda]$ . Taking the expectation over the drawing of  $\text{MOCKRUN}$ , we get that  $\mathbb{E}[c(\text{OPT}(\text{MOCKRUN}))] \leq Z$ . Since algorithm  $\mathcal{A}$  is  $\Delta$ -competitive, we immediately get that  $\mathbb{E}[c(z_0)] \leq \Delta \cdot Z$ .

It remains to bound the cost of  $z_1$ . This second half of the proof resembles that of Theorem 3.1. Let  $z_0(v)$  denote the state of the solution held by  $\mathcal{A}$  before the arrival of request  $v$ .

Fix an index  $t \in [n]$ . Without loss of generality, reorder the corresponding  $\text{MOCKRUN}$  samples  $\widehat{v}_1^t, \dots, \widehat{v}_\lambda^t$  to agree with their relative order in  $\pi$ . Define  $\text{MOCKRUN}_{<v}$  to be the set of clients of  $\text{MOCKRUN}$  that arrive before  $v$  according to order  $\pi$ . By Definition 2.1, since  $z_0(\widehat{v}_1^t) \leq \dots \leq z_0(v^t)$  for every realization of the random variables, we have that for each request  $v$ ,

$$\begin{aligned} \text{AUG} \left( v \mid z_0(\widehat{v}_1^t), \text{MOCKRUN}_{<\widehat{v}_1^t} \right) &\geq \dots \geq \text{AUG} \left( v \mid z_0(\widehat{v}_\lambda^t), \text{MOCKRUN}_{<\widehat{v}_\lambda^t} \right) \\ &\geq \text{AUG} \left( v \mid z_0(v^t), \text{MOCKRUN} \cup \{v^1, \dots, v^{t-1}\} \right). \end{aligned}$$

Then, taking the expectation over both the random sequences of  $z_0(\widehat{v}_1^t), \dots, z_0(v^t)$  and the identically distributed draws of  $\widehat{v}_1^t, \dots, \widehat{v}_\lambda^t, v^t \sim D^t$ ,

$$\begin{aligned} \mathbb{E} \left[ \text{AUG} \left( \widehat{v}_1^t \mid z_0(\widehat{v}_1^t), \text{MOCKRUN}_{<\widehat{v}_1^t} \right) \right] &\geq \dots \geq \mathbb{E} \left[ \text{AUG} \left( \widehat{v}_\lambda^t \mid z_0(\widehat{v}_\lambda^t), \text{MOCKRUN}_{<\widehat{v}_\lambda^t} \right) \right] \\ &\geq \mathbb{E} \left[ \text{AUG} \left( v^t \mid z_0(v^t), \text{MOCKRUN} \cup \{v^1, \dots, v^{t-1}\} \right) \right]. \end{aligned}$$

Summing yields

$$\lambda \cdot \mathbb{E} \left[ \text{AUG} \left( v^t \mid z_0(v^t), \text{MOCKRUN} \cup \{v^1, \dots, v^{t-1}\} \right) \right] \leq \sum_{i=1}^{\lambda} \mathbb{E} \left[ \text{AUG} \left( \widehat{v}_i^t \mid z_0(\widehat{v}_i^t), \text{MOCKRUN}_{<\widehat{v}_i^t} \right) \right].$$

Finally, summing again over  $t \in [n]$ , we get that the second-stage costs of Algorithm 2 are bounded by

$$\begin{aligned} \lambda \cdot \mathbb{E}[c(z_1)] &= \sum_{t=1}^n \lambda \cdot \mathbb{E}[\text{AUG}(v^t \mid z_0(v^t), \text{MOCKRUN} \cup \{v^1, \dots, v^{t-1}\})] \\ &\leq \sum_{t=1}^n \sum_{i=1}^{\lambda} \mathbb{E}[\text{AUG}(\widehat{v}_i^t \mid z_0(\widehat{v}_i^t), \text{MOCKRUN}_{<\widehat{v}_i^t})] \\ &\leq \mathbb{E}[c(z_0)], \end{aligned}$$

where the last inequality holds since  $\mathcal{A}$  pays at least  $\text{AUG}(\widehat{v}_i^t \mid z_0(\widehat{v}_i^t), \text{MOCKRUN}_{<\widehat{v}_i^t})$  in rounds where  $\widehat{v}_i^t$  is unsatisfied on arrival.

We conclude that the total expenditure of the algorithm is  $\mathbb{E}[c(z_0) + \lambda \cdot c(z_1)] \leq 2\mathbb{E}[c(z_0)] \leq 2\Delta \cdot Z$ .  $\square$

## 5 Online-With-a-Sample

In this section we show a general reduction from online-with-a-sample AIPs to the random-order version. Recall that in this model,  $\alpha$  is the proportion of the input sequence which is uniformly randomly chosen to be revealed to the algorithm up front.

**THEOREM 5.1.** *Let  $\mathcal{I}$  be an instance class of AIPs. If algorithm  $\mathcal{A}$  is a random-order AIP algorithm with competitive ratio  $\Delta$  on class  $\mathcal{I}$ , then there is an online-with-a-sample algorithm  $\mathcal{A}'$  for class  $\mathcal{I}$  with competitive ratio  $O(\Delta/\alpha)$ .*

*Proof.* Let  $\mathcal{A}$  be the random-order algorithm for instance class  $\mathcal{I}$  with expected competitive ratio  $\Delta$ . Define  $\mathcal{A}'$  to be Algorithm 3.

---

### Algorithm 3 ONLINE-WITH-A-SAMPLE TO RANDOM ORDER

---

- 1: Let  $\text{SAMPLES} = \{\widehat{v}^1, \dots, \widehat{v}^{\alpha \cdot n}\}$  be the samples given upfront.
  - 2: **for**  $\tau = 1, 2, \dots, \alpha \cdot n$  **do**
  - 3:      $\pi^\tau \leftarrow$  the  $\tau^{\text{th}}$  sample in random order.
  - 4:     Feed  $\pi^\tau$  to  $\mathcal{A}$ .
  - 5: Let  $\widehat{z}$  be the output of  $\mathcal{A}$ .
  - 6: Initialize  $z \leftarrow \widehat{z}$ .
  - 7: **for**  $t = 1, 2, \dots, n$  **do**
  - 8:     Draw  $v^t \sim D^t$ .
  - 9:     **if**  $v^t$  not satisfied by  $z$  **then**
  - 10:         Update  $z \leftarrow \max(z, \text{BACKUP}(v^t \mid z, \text{SAMPLES} \cup \{v^1, \dots, v^{t-1}\}))$ .
  - 11: **return**  $z$ .
- 

Assume without loss of generality that  $n$  is a multiple of  $1/\alpha$ . Further assume that  $1/\alpha \in \mathbb{N}$ ; this can be done by taking  $\alpha \leftarrow 1/\lceil 1/\alpha \rceil$  and paying at most a factor of two.

We imagine generating  $\text{SAMPLES}$  according to the following procedure. First, perform a random partition of  $v^1, \dots, v^n$  into  $\alpha \cdot n$  groups of size  $1/\alpha$ . Pick a uniformly random representative request from each subset to include in  $\text{SAMPLES}$ . For each request  $v$ , let  $r(v)$  be the representative request of the group containing  $v$ , and let  $\text{SAMPLES}_{<v}$  be the set of samples that arrive before  $v$  according to random order  $\pi$ . Now:

$$\begin{aligned} \mathbb{E}[\text{AUG}(v^t \mid z(v^t), \text{SAMPLES} \cup \{v^1, \dots, v^{t-1}\})] &\leq \mathbb{E}[\text{AUG}(v^t \mid \widehat{z}(r(v^t)), \text{SAMPLES}_{<r(v^t)})] \\ &= \mathbb{E}[\text{AUG}(r(v^t) \mid \widehat{z}(r(v^t)), \text{SAMPLES}_{<r(v^t)})]. \end{aligned}$$



The inequality holds by the monotonicity of augmentation condition of Definition 2.1, since  $z(v^t) \geq \widehat{z}(r(v^t))$ . The equality holds because  $v^t$  and  $r(v^t)$  are identically distributed at the beginning of the run. Summing over  $t$ , and noting that  $\mathcal{A}$  must pay at least  $\text{AUG}(\widehat{v}^\tau \mid \widehat{z}(v^\tau))$  in round  $\tau$  when  $\widehat{v}^\tau$  is unsatisfied on arrival, we get that the total backup cost is bounded as

$$\begin{aligned} \sum_{t=1}^n \mathbb{E}[\text{AUG}(v^t \mid z(v^t), \text{SAMPLES} \cup \{v^1, \dots, v^{t-1}\})] &\leq \sum_{t=1}^n \mathbb{E}[\text{AUG}(r(v^t) \mid z(r(v^t)), \text{SAMPLES}_{<r(v^t)})] \\ &= \sum_{\tau=1}^{\alpha \cdot n} \frac{1}{\alpha} \cdot \mathbb{E}[\text{AUG}(\widehat{v}^\tau \mid z(\widehat{v}^\tau), \text{SAMPLES}_{<\widehat{v}^\tau})] \\ &\leq \frac{1}{\alpha} \mathbb{E}[c(\widehat{z})]. \end{aligned}$$

The equality above comes from the fact that each representative that makes up  $\text{SAMPLES}$  appears  $1/\alpha$  times in the first sum. We can bound the expression above by  $\Delta/\alpha \cdot \mathbb{E}[c(\text{OPT}(v_1, \dots, v_n))]$  by the assumption on  $\mathcal{A}$ , and so in total, Algorithm 1 pays at most  $\Delta/\alpha \cdot \mathbb{E}[c(\text{OPT})]$ .  $\square$

We conclude by remarking that a similar argument proves Theorem 5.1 in the slightly different ‘online-with-a-sample’ setting in which each of the requests in the online sequence is sampled independently with probability  $\alpha$ . More generally, for  $\alpha \leq 1$  it is possible to get an analogous tradeoff between  $\alpha$  in the number of samples and  $1/\alpha$  in the approximation ratio for Theorem 3.1 and Theorem 4.1 as well.

## 6 Online Facility Location

In this section we apply our framework to online facility location problems.

**6.1 Facility Location** In `FACILITYLOCATION`, the input is a set of  $n$  clients and  $m$  facilities. Each facility  $f$  has an opening cost  $c_f$ , and each client-facility pair  $(v, f)$  has a connection cost  $c_{fv}$ . The goal is to open a number of facilities and connect each client to exactly one open facility such that the total cost is minimized. A standard integer programming formulation for offline (unit-demand) `FACILITYLOCATION` is as follows [CCZ14]:

$$(6.1) \quad \begin{aligned} \min \quad & \sum_f c_f \cdot x_f + \sum_{f,v} c_{fv} \cdot y_{fv} \\ & \sum_j y_{fv} \geq 1 \quad \forall v \\ & y_{fv} \leq x_f \quad \forall f, v \\ & x_f, y_{fv} \in \{0, 1\}. \end{aligned}$$

In the random-order online version, the  $m$  facilities are known ahead of time, and  $n$  unknown clients arrive online in random order; on the arrival of each client, the algorithm must choose which (if any) new facilities to open, and then connect the client to an open facility. Decisions are irrevocable, in the sense that a client may not change which facility it has connected to after arrival.

We observe that `FACILITYLOCATION` is amenable to our framework:

**OBSERVATION 6.1.** *FACILITYLOCATION is an AIP.*

This enables us to convert algorithms for random-order `FACILITYLOCATION` into algorithms for the prophet, two-stage, and with-a-sample settings.

**6.2 Metric Facility Location** Random-order `FACILITYLOCATION` is well studied when the connection costs  $c_{fv}$  satisfy the triangle inequality. In pioneering work, Meyerson gave an 8-approximation for this problem [Mey01], which has recently been improved to a 3-approximation [KNR23]. Appealing to Lemma 6.1 and plugging this algorithm as a black box into Theorems 3.1, 4.1, and 5.1, we get:

**COROLLARY 6.1.** *For the METRICFACILITYLOCATION problem, there exists a 6-competitive algorithm in the single-sample prophet setting, a 6-competitive algorithm in the 2-stage prophet setting, and a  $O(1/\alpha)$ -competitive algorithm in the online-with-a-sample setting.*

In the next section, we study the more general problem without the metric assumption.

**6.3 Non-Metric Facility Location** When connection costs do not satisfy the triangle inequality, this problem is more complex. In particular, NONMETRICFACILITYLOCATION recovers SETCOVER as the special case in which all service costs are  $c_{fv} \in \{0, \infty\}$ . In this section, we give an algorithm for random-order NONMETRICFACILITYLOCATION that is best possible, even in the special case of set cover. We show:

**THEOREM 6.1.** *There exists an  $O(\log mn)$ -competitive algorithm for random-order NONMETRICFACILITYLOCATION.*

From Lemma 6.1 and Theorems 3.1, 4.1, and 5.1 we then directly obtain:

**COROLLARY 6.2.** *For random-order NONMETRICFACILITYLOCATION, there exists an  $O(\log mn)$ -competitive algorithm in the single-sample prophet setting, a  $O(\log mn)$ -competitive algorithm in the 2-stage prophet setting, and a  $O(\log mn/\alpha)$ -competitive algorithm in the online-with-a-sample setting.*

We demonstrate the versatility of the the LEARNORCOVER algorithm of [GKL21] for random-order set cover by adapting it to random-order FACILITYLOCATION, which is not a pure covering problem. The challenge is to decide which facilities to open in any given round; once this is decided, one can assume the incoming client in that round always connects to the cheapest open facility.

Let  $\mathcal{C}^t$  be the facilities purchased by the end of round  $t$ . For every client  $v$ , define

$$f^t(v) \in \operatorname{argmin}_f (\mathbf{1}\{f \notin \mathcal{C}^t\} \cdot c_f + c_{fv})$$

$$\kappa_v^t := \min_f (\mathbf{1}\{f \notin \mathcal{C}^t\} \cdot c_f + c_{fv})$$

to be the facility which can connect the client in the cheapest way possible (including opening said facility if necessary), and the corresponding marginal cost of doing so. This is the cost at the end of round  $t$ . Note that  $\kappa_v^t$  corresponds to  $\text{AUG}(v^t \mid \mathcal{C}^t, \{v^1, \dots, v^{t-1}\})$  in our more general notation. Also define

$$\Gamma^t(v) := \{f : c_{fv} \leq \kappa_v^t/2\}$$

to be the set of set of facilities that, if opened, would reduce the marginal cost of connecting  $v$  by at least a factor of 2. Then we will say a facility  $f$  covers a client  $v$  at time  $t$  if  $f \in \Gamma^t(v)$ .

We will show that Algorithm 4 is in expectation an  $O(\log mn)$ -approximation for random-order NONMETRICFACILITYLOCATION. Our approach may be viewed as running LEARNORCOVER algorithm for set cover, but on a *dynamically changing* set system. The facilities are the sets, the clients are the elements, and a client's element is contained in a facility's set at time  $t$  if  $f \in \Gamma^t(v)$ .

By a guess-and-double approach, we may assume the algorithm has access to a bound  $\beta$  such that  $\text{LP}_{\text{OPT}} \leq \beta \leq 2 \cdot \text{LP}_{\text{OPT}}$ ; here  $\text{LP}_{\text{OPT}}$  is the cost of the optimal solution for the linear programming relaxation of (6.1) for the given NONMETRICFACILITYLOCATION instance. We will denote by  $\Xi^t$  the event that the arriving client  $v^t$  satisfies the condition on Line 5 that  $\kappa_{v^t}^{t-1} \geq \beta/t$  and the bulk of Algorithm 4 is executed. We will say that  $v^t$  is *preemptively connected* if this condition is not met; this is the event  $\neg \Xi^t$ .

Through Line 10 we maintain:

**INVARIANT 6.1.** *For all time steps  $t$ , it holds that  $\langle c, x^t \rangle = \beta$ .*

We start by defining notation. Let  $\text{OPT} = (x^*, y^*)$  be an optimal fractional solution. Let  $U^t = \{v^{t+1}, \dots, v^n\}$  be the clients remaining uncovered at the end of round  $t$  (where  $U^0 = U$  is the entire client set). Let  $X^t(v) := \sum_{f \in \Gamma^{t-1}(v)} x_f^{t-1}$  be the fractional weight of facilities  $f$  which cover  $v$  at time  $t - 1$ . We define  $\rho^t := \sum_v \kappa_v^t$ , and consider the following potential:

$$\Phi(t) := C_1 \cdot \underbrace{\left( \text{KL}_c(x^* \parallel x^t) + 2 \cdot \sum_{v \in U^t} \sum_f c_{fv} \cdot y_{fv}^* \right)}_{\Phi_L(t)} + C_2 \cdot \underbrace{\beta \cdot \log \left( \frac{\rho^t}{\beta} + \frac{1}{n} \right)}_{\Phi_C(t)},$$

---

**Algorithm 4** LEARNORCOVERNMFL
 

---

```

1: Let  $\mathcal{F}' \leftarrow \{f : \beta/m \leq c_f \leq \beta\}$  and let  $m' \leftarrow |\mathcal{F}'|$ .
2: Initialize  $x_f^0 \leftarrow \frac{\beta}{c_f \cdot m'} \cdot \mathbf{1}\{f \in \mathcal{F}'\}$ .
3: for  $t = 1, 2, \dots, n$  do
4:    $v^t \leftarrow t^{\text{th}}$  client in the random order, and let  $\mathcal{R}^t \leftarrow \emptyset$ .
5:   if  $\kappa_{v^t}^{t-1} \geq \beta/t$  then
6:     for each facility  $f$ , add  $\mathcal{R}^t \leftarrow \mathcal{R}^t \cup \{f\}$  with probability  $\min(\kappa_{v^t}^{t-1} \cdot x_f^{t-1}/\beta, 1)$ .
7:     Update  $\mathcal{C}^t \leftarrow \mathcal{C}^{t-1} \cup \mathcal{R}^t$ .
8:     if  $\sum_{f \in \Gamma^{t-1}(v^t)} x_f^{t-1} < 1$  then
9:       For every facility  $f$ , update  $x_f^t \leftarrow x_f^{t-1} \cdot \exp\{\mathbf{1}\{f \in \Gamma^{t-1}(v^t)\} \cdot \kappa_{v^t}^{t-1}/c_f\}$ .
10:      Let  $Z^t = \langle c, x^t \rangle / \beta$  and normalize  $x^t \leftarrow x^t / Z^t$ .
11:     else
12:        $x^t \leftarrow x^{t-1}$ .
13:   else
14:      $x^t \leftarrow x^{t-1}$ .
15:    $\mathcal{C}^t \leftarrow \mathcal{C}^t \cup \{f^{t-1}(v^t)\}$  and connect  $v^t$  to  $f^{t-1}(v^t)$ .
16: return  $\mathcal{C}$ .
    
```

---

where the constants  $C_1$  and  $C_2$  will be determined later. We will refer to  $\Phi_L$  as the “learning” portion of the potential and  $\Phi_C$  as the “covering” portion of the potential.

Our potential  $\Phi(t)$  resembles the one used to analyze LEARNORCOVER for set cover [GKL21] which can also be decomposed into “learning” and “covering” portions. Their learning portion also involves a KL-divergence term, but ours is more intricate since we additionally charge to the connection cost paid by fractional OPT.

LEMMA 6.1. (BOUNDS ON  $\Phi$ ) *The initial potential is bounded as  $\Phi(0) = O(\beta \cdot \log mn)$ , and  $\Phi(t) \geq -\beta \cdot \log n$  for all  $t$ .*

We now show the potential decreases sufficiently in every round. We bound the decrease of each term in the potential separately.

LEMMA 6.2. (CHANGE IN  $\Phi_L$ ) *For rounds when the event  $\Xi^t$  holds, the expected change in  $\Phi_L$  is*

$$\begin{aligned}
 & \mathbb{E}_{v^t, \mathcal{R}^t} [\Phi_L(t) - \Phi_L(t-1) \mid x^{t-1}, U^{t-1}, \Xi^t] \\
 (6.2) \quad & \leq \mathbb{E}_{v \sim U^{t-1}} \left[ \frac{e^2 - 1}{2} \cdot \kappa_v^{t-1} \cdot \min(X^t(v), 1) - \kappa_v^{t-1} \right].
 \end{aligned}$$

When the arriving  $v^t$  is preemptively connected and  $\Xi^t$  does not hold,

$$(6.3) \quad \Phi_L(t) - \Phi_L(t-1) \leq 0.$$

Note that the expected change in the statement above depends only on the randomness of the arriving uncovered client  $v^t$ , not on the randomly chosen facilities  $\mathcal{R}^t$ . On the other hand we can bound the change in  $\Phi_C$  as follows.

LEMMA 6.3. (CHANGE IN  $\Phi_C$ ) *For all rounds  $t$  for which  $\Xi^t$  holds, the expected change in  $\Phi_C$  is*

$$(6.4) \quad \mathbb{E}_{v^t, \mathcal{R}^t} [\Phi_C(t) - \Phi_C(t-1) \mid x^{t-1}, U^{t-1}, \Xi^t] \leq -\frac{1 - e^{-1}}{4} \cdot \mathbb{E}_{u \sim U^{t-1}} [\kappa_u^{t-1} \cdot \min(X^t(v), 1)].$$

For rounds in which  $\Xi^t$  does not hold,

$$(6.5) \quad \Phi_C(t) - \Phi_C(t-1) \leq 0.$$

We defer the proof of Lemmas 6.1, 6.2, and 6.3 to Appendix A, and we now show how to combine them to prove the theorem.

*Proof of Theorem 6.1.* Let  $c(\text{ALG}(t))$  be the cost paid by Algorithm 4 up to and including time  $t$ , and furthermore let  $c_{\text{PRE}}(\text{ALG}(t))$  denote the cost paid by Algorithm 4 for clients which are preemptively connected on Line 15 up to and including round  $t$ , and  $c_{\text{LOC}}(\text{ALG}(t))$  denote the cost paid by Algorithm 4 on Lines 6 and 15 up to and including round  $t$  during rounds in which  $v^t$  is not preemptively connected and  $\Xi^t$  holds. (PRE is for preemptive, LOC is for LEARNORCOVER).

We can provide a simple bound on the cheap facility-client connections which the algorithm buys on Line 15 in the event that  $v^t$  is preemptively connected (that is, in the event that  $\Xi^t$  does not hold). These connections collectively cost at most

$$\begin{aligned}
 c_{\text{PRE}}(\text{ALG}(n)) &= c_{\text{PRE}}(\text{ALG}(n)) - c_{\text{PRE}}(\text{ALG}(0)) \\
 &= \sum_t (c_{\text{PRE}}(\text{ALG}(t)) - c_{\text{PRE}}(\text{ALG}(t-1))) \\
 (6.6) \quad &\leq \sum_{t \in [n]} \frac{\beta}{t} = O(\beta \cdot \log n).
 \end{aligned}$$

We now consider the per-round costs incurred by the bulk of the algorithm, during the rounds in which  $\Xi^t$  holds. In every round  $t$ , the expected cost of the sampled facilities  $\mathcal{R}^t$  in Line 6 is  $\kappa_{v^t}^{t-1} \cdot \langle c, x^{t-1} \rangle / \beta = \kappa_{v^t}^{t-1}$  (by Invariant 6.1). The algorithm pays at most an additional  $\kappa_{v^t}^{t-1}$  in Line 15, and hence the total expected cost per round is at most  $2 \cdot \kappa_{v^t}^{t-1}$ .

By combining Lemmas 6.2 and 6.3, and setting the constants  $C_1 = 2$  and  $C_2 = 4e(e+1)$ , we have

$$\begin{aligned}
 &\mathbb{E}_{v^t, \mathcal{R}^t} [\Phi(t) - \Phi(t-1) \mid v^1, \dots, v^{t-1}, \mathcal{R}^1, \dots, \mathcal{R}^{t-1}, \Xi^t] \\
 &= \mathbb{E}_{v^t, \mathcal{R}^t} \left[ \begin{array}{c} C_1 \cdot (\Phi_L(t) - \Phi_L(t-1)) \\ + C_2 \cdot (\Phi_C(t) - \Phi_C(t-1)) \end{array} \mid v^1, \dots, v^{t-1}, \mathcal{R}^1, \dots, \mathcal{R}^{t-1}, \Xi^t \right] \\
 &\leq - \mathbb{E}_{v^t, \mathcal{R}^t} [2 \cdot \kappa_{v^t}^{t-1} \mid v^1, \dots, v^{t-1}, \mathcal{R}^1, \dots, \mathcal{R}^{t-1}, \Xi^t],
 \end{aligned}$$

which cancels the expected change in  $c_{\text{LOC}}$  in each round. We therefore have the inequality

$$(6.7) \quad \mathbb{E}_{v^t, \mathcal{R}^t} [\Phi(t) - \Phi(t-1) + c_{\text{LOC}}(\text{ALG}(t)) - c_{\text{LOC}}(\text{ALG}(t-1)) \mid v^1, \dots, v^{t-1}, \mathcal{R}^1, \dots, \mathcal{R}^{t-1}] \leq 0,$$

where we used that the change in  $\Phi_L$  and  $\Phi_C$  is at most 0 for rounds in which  $\Xi^t$  does not hold.

By repeatedly applying (6.7) for all  $1 \leq t \leq n$ , we obtain

$$\begin{aligned}
 (6.8) \quad \mathbb{E}_{v, \mathcal{R}} [\Phi(n) - \Phi(0) + c_{\text{LOC}}(\text{ALG}(n)) - c_{\text{LOC}}(\text{ALG}(0))] &\leq 0 \\
 \mathbb{E}_{v, \mathcal{R}} [c_{\text{LOC}}(\text{ALG}(n))] &\leq \Phi(0) + c_{\text{LOC}}(\text{ALG}(0)) - \mathbb{E}_{v, \mathcal{R}} [\Phi(n)] \\
 &\leq O(\beta \cdot \log mn) + 0 - (-\beta \cdot \log n),
 \end{aligned}$$

where (6.8) follows from the fact that  $c_{\text{LOC}}(\text{ALG}(0)) = 0$ , together with the bounds on  $\Phi$  established in Lemma 6.1.

To conclude, by (6.6) and (6.8) we have

$$\begin{aligned}
 (6.9) \quad \mathbb{E}_{v, \mathcal{R}} [c(\text{ALG}(n))] &= \mathbb{E}_{v, \mathcal{R}} [c_{\text{PRE}}(\text{ALG}(n))] + \mathbb{E}_{v, \mathcal{R}} [c_{\text{LOC}}(\text{ALG}(n))] \\
 &\leq \beta + O(\beta \cdot \log n) + O(\beta \cdot \log mn),
 \end{aligned}$$

as desired.  $\square$

This concludes our discussion of random-order FACILITYLOCATION. Our results settle the approximability of both the metric and non-metric versions this problem in the random-order model, and hence also in the prophet, 2-stage prophet, and online-with-a-sample models.

## 7 Set Multicover with Multiplicity Constraints

We now give a second application of our framework to set multicover with multiplicity constraints, which we refer to as SETMULTICOVER. This generalizes unit-cost set cover and additionally introduces non-covering constraints; in particular, multiplicity constraints on the decision variables. The set multicover problem can be formally written as the following IP:

$$(7.1) \quad \begin{aligned} \min_z \quad & \langle 1, z \rangle \\ \text{s.t.} \quad & Az \geq b \\ & z \in \{0, 1\}^m, \end{aligned}$$

where the entries of  $A$  are in  $\{0, 1\}$ . It is then without loss of generality to assume that  $b \in \mathbb{Z}_+^n$ . We recover SETCOVER as the special case where  $b = 1$ . We again observe that SETMULTICOVER is amenable to our framework:

OBSERVATION 7.1. *SETMULTICOVER is an AIP.*

This enables us to convert algorithms for random-order SETMULTICOVER into algorithms for the prophet, two-stage, and with-a-sample settings.

In random-order SETMULTICOVER the rows of  $A$  are revealed in random-order, and the algorithm must maintain a monotonically increasing solution  $z$  that satisfies all constraints revealed thus far. We show:

THEOREM 7.1. *There exists an  $O(\log mn)$ -competitive algorithm for random-order SETMULTICOVER.*

By appealing to Lemma 7.1 and Theorems 3.1, 4.1, and 5.1 we then directly obtain:

COROLLARY 7.1. *For SETMULTICOVER, there exists an  $O(\log mn)$ -competitive algorithm in the single-sample prophet setting, a  $O(\log mn)$ -competitive algorithm in the 2-stage prophet setting, and a  $O(\log(mn)/\alpha)$ -competitive algorithm in the online-with-a-sample setting.*

Once again we show how to adapt ideas from [GKL21] to this more general setting. This time we extend their more general algorithm for random-order CIPs. We will make use of the preliminaries given in Section 2, and our approach will be similar to that of Section 6.

Let  $z^t$  denote the integer solution in round  $t$ . Let  $d_i^t$  denote the undercoverage of  $i$  in the beginning of round  $t$ ; that is  $d_i^t := \max(0, b_i - \langle a_i, z^{t-1} \rangle)$  where  $z^t$  is the integer solution at the end of round  $t$ . Now we define

$$\rho^t := \sum_{i \in [n]} d_i^t.$$

We pursue a guess-and-double approach to identifying  $\beta$  such that  $\text{LP}_{\text{OPT}} \leq \beta \leq 2 \cdot \text{LP}_{\text{OPT}}$ , where  $\text{LP}_{\text{OPT}}$  is the cost of an optimal fractional solution  $x^*$  to (7.1). We will maintain a solution  $x^t$  to (7.1) of cost  $\beta$ .

In a manner similar to Algorithm 4, our algorithm will react differently to constraints which arrive uncovered. In Algorithm 4 the criterion was that arriving clients cannot be cheaply connectable; here we perform a LEARNORCOVER step if the element arrives undercovered, meaning that  $d_{i_t}^t > 0$  on Line 4.

THEOREM 7.2. *For set multicover with multiplicity constraints, Algorithm 5 achieves an expected competitive ratio of  $O(\log mn)$ .*

Our potential is

$$\Phi(t) := C_1 \cdot \underbrace{\text{KL}(x^* \parallel x^t)}_{\Phi_L(t)} + C_2 \cdot \underbrace{\beta \cdot \log\left(\frac{\rho^t}{\beta} + \frac{1}{m}\right)}_{\Phi_C(t)},$$

where again we view  $\Phi_L$  is the learning part of the potential and  $\Phi_C$  is the covering part. We will fix constants  $C_1$  and  $C_2$  later. To begin, we bound the value of this potential:

---

**Algorithm 5** LEARNORCOVERSMC
 

---

```

1: Initialize  $x_j^0 \leftarrow \frac{\beta}{m}$  for every  $j$ , and set  $z_j^0 \leftarrow 0$ .
2: for  $t = 1, 2, \dots$  do
3:    $i \leftarrow t$ -th constraint in the random order.
4:   if  $i$  is not covered on arrival then
5:     Let  $\mathcal{T}^t := \{j : z_j^{t-1} = 0, a_{ij} = 1\}$ . // unbought coordinates covering  $i$ 
6:     Let  $X_i^{t-1} := \sum_{j \in \mathcal{T}^t} x_j^{t-1}$ . // frac. coverage by unbought  $j$ 
7:      $d_i^t := b_i - \langle a_i, z^{t-1} \rangle$ . // integral uncoverage
8:     for each  $j$ , sample  $z_j^t \leftarrow \text{Ber}(d_i^t \cdot x_j^{t-1} / \beta)$ .
9:     if  $X_i^{t-1} \leq d_i^t$  then //  $i$  fractionally undercovered
10:      Set  $z_j^t \leftarrow 1$  for  $j \in \mathcal{T}^t$  with  $x_j^{t-1} \geq 1/e$ .
11:      for each  $j$ , update  $x_j^t \leftarrow x_j^{t-1} \cdot \exp\{\mathbf{1}\{j \in \mathcal{T}^t\}\}$ .
12:      Let  $Z^{(t)} := \langle \mathbf{1}, x^t \rangle / \beta$  and renormalize  $x^t \leftarrow x^t / Z^{(t)}$ .
13:     if  $i$  still uncovered then
14:        $x^t \leftarrow x^{t-1}$  and  $z^t \leftarrow z^{t-1}$ .
15:        $z_j^t \leftarrow 1$  for  $d_i^t$ -many arbitrary  $j \in \mathcal{T}^t$ . // buy backup sets
16:     else
17:        $x^t \leftarrow x^{t-1}$  and  $z^t \leftarrow z^{t-1}$ .
18: return  $z$ 
    
```

---

LEMMA 7.1. (BOUNDS ON  $\Phi$ ) *The initial potential is bounded as  $\Phi(0) = O(\beta \cdot \log mn)$ , and  $\Phi(t) \geq -\beta \cdot \log(n)$  for all rounds  $t$ .*

Let  $U^t$  denote the constraints  $i$  which have not yet arrived in round  $t$ ; that is,  $U^t = \{i^t, i^{t+1}, \dots, i^n\}$ . We now turn to the expected change in the learning portion of the potential in each round.

LEMMA 7.2. (CHANGE IN  $\Phi_L$ ) *For rounds in which  $i^t$  arrives uncovered, the expected change in  $\Phi_L$  is*

$$(7.2) \quad \mathbb{E}_{i^t, \mathcal{R}^t} [\Phi_L(t) - \Phi_L(t-1) \mid x^{t-1}, U^{t-1}, d_{i^t}^t > 0] \leq \mathbb{E}_{i \sim U^{t-1}} [(e-1) \cdot \min(X_i^{t-1}, d_i^t) - d_i^t].$$

When  $i^t$  is covered on arrival,

$$(7.3) \quad \Phi_L(t) - \Phi_L(t-1) \leq 0.$$

We next bound the expected change in  $\Phi_C(t)$ .

LEMMA 7.3. (CHANGE IN  $\Phi_C$ ) *In every round for which  $i^t$  is uncovered on arrival, the expected change in  $\Phi_C$  is*

$$(7.4) \quad \mathbb{E}_{i^t, \mathcal{R}^t} [\Phi_C(t) - \Phi_C(t-1) \mid x^{t-1}, U^{t-1}, d_{i^t}^t > 0] \leq -\frac{\gamma}{2} \cdot \mathbb{E}_{i \sim U^{t-1}} [\min(X_i^{t-1}, d_i^t)],$$

where  $\gamma$  is a fixed constant. On rounds in which  $i^t$  arrives covered,

$$(7.5) \quad \Phi_C(t) - \Phi_C(t-1) \leq 0.$$

We again defer the proofs of Lemmas 7.1, 7.2, and 7.3 to Section A, and now show how to combine them to bound the expected cost of Algorithm 5:

*Proof of Theorem 7.2.* Combining Lemma 7.2 and Lemma 7.3, choosing  $C_1 = (e+2)$  and  $C_2 = 2 \cdot (e+2)(e-1)/\gamma$ , and recalling that  $d_{i^t}^t = 0$  in rounds for which  $i^t$  arrives covered, we have that

$$\mathbb{E}_{i^t, \mathcal{R}^t} [\Phi(t) - \Phi(t-1) \mid i^1, \dots, i^{t-1}, \mathcal{R}^1, \dots, \mathcal{R}^{t-1}]$$

$$\begin{aligned}
 &= \mathbb{E}_{i^t, \mathcal{R}^t} \left[ \begin{array}{c} C_1 \cdot (\Phi_L(t) - \Phi_L(t-1)) \\ + C_2 \cdot (\Phi_C(t) - \Phi_C(t-1)) \end{array} \middle| i^1, \dots, i^{t-1}, \mathcal{R}^1, \dots, \mathcal{R}^{t-1} \right] \\
 (7.6) \quad &\leq - \mathbb{E}_{i^t, \mathcal{R}^t} [(e+2) \cdot d_i^t \mid i^1, \dots, i^{t-1}, \mathcal{R}^1, \dots, \mathcal{R}^{t-1}]
 \end{aligned}$$

for all rounds  $t$ . In each round the algorithm buys at most  $e \cdot d_i^t$  coordinates in Line 10 (since this only happens in the case when  $X_i^{t-1} \leq d_i^t$ ), and samples  $d_i^t$  sets in expectation in Line 8, and buys at most  $d_i^t$  coordinates in Line 15, for a total of at most  $(e+2) \cdot d_i^t$  sets bought in expectation. From (7.6) we therefore have that for all rounds  $t$ ,

$$(7.7) \quad \mathbb{E}_{\substack{i^t \sim U^{(t)} \\ \mathcal{R} \sim x^{(t)}}} \left[ \Phi(t) - \Phi(t-1) + c(\text{ALG}(t)) - c(\text{ALG}(t-1)) \mid i^{(1)}, \dots, i^{(t-1)}, \mathcal{R}^{(1)}, \dots, \mathcal{R}^{(t-1)} \right] \leq 0.$$

Repeatedly applying (7.7) for all  $1 \leq t \leq n$  yields

$$\begin{aligned}
 \mathbb{E}_{i, \mathcal{R}} [\Phi(n) - \Phi(0) + c(\text{ALG}(n)) - c(\text{ALG}(0))] &\leq 0 \\
 \mathbb{E}_{i, \mathcal{R}} [c(\text{ALG}(n))] &\leq c(\text{ALG}(0)) + \Phi(0) - \mathbb{E}_{i, \mathcal{R}} [\Phi(n)],
 \end{aligned}$$

and so observing that  $c(\text{ALG}(0)) = 0$  and applying Line 7.1 we have

$$(7.8) \quad \mathbb{E}_{i, \mathcal{R}} [c(\text{ALG}(n))] \leq O(\beta \cdot \log mn),$$

as desired.  $\square$

This concludes our discussion of SETMULTICOVER. Once again our results settle the approximability of this problem in the random-order, prophet, 2-stage prophet, and online-with-a-sample models. This gives a partial answer towards the question of [GKL21] on whether the same results are possible for general box-constrained CIPs; it remains a tantalizing open question to understand the general case.

## 8 Conclusion

In this paper we showed that stochastic set cover can be solved “even more obviously” than [GGL<sup>+</sup>13], with only coarse advice about the process generating the input. It is tempting to try to relax these online-with-advice models for set cover by allowing for bounded error in the advice. We discuss why this is challenging in Appendix B, as some natural candidates for relaxed models have strong lower bounds.

We submit as an interesting open problem the task of determining the tight dependence on  $\alpha$  in Theorem 1.3. We conjecture that it should be  $O(\log(mn) \log(1/\alpha))$ . Theorem 1.3 implies that when  $\alpha = \Theta(1)$  there is an  $O(\log mn)$  competitive algorithm, and when  $\alpha = 1/\text{poly}(n)$  the  $O(\log m \log n)$  competitive algorithm of [BN09] is best possible; this conjecture interpolates smoothly between these extremes.

**Acknowledgements** Roie Levin would like to thank Guy Even for asking about the 2-stage prophet model, and Niv Buchbinder for helpful discussions.

## A Deferred Proofs

Here we present the proofs of lemmas supporting our random-order algorithms of Sections 6 and 7.

### A.1 Facility Location

LEMMA 6.1. (BOUNDS ON  $\Phi$ ) *The initial potential is bounded as  $\Phi(0) = O(\beta \cdot \log mn)$ , and  $\Phi(t) \geq -\beta \cdot \log n$  for all  $t$ .*

*Proof.* To begin, we claim that for all  $f$  in the support of  $x^*$ , we have  $c_f \leq \beta$ . To see this, consider an optimal fractional solution  $(x^*, y^*)$  to (6.1), and assume for the sake of contradiction that there is some facility  $\hat{f} \in \text{support}(x^*)$  for which  $x_{\hat{f}}^* > \beta$ . This  $\hat{f}$  provides some fractional connection to some clients; let  $\epsilon := \min(y_{\hat{f},v}^* : y_{\hat{f},v}^* > 0)$  be the minimum connection provided to all such clients. Finally, consider the perturbed

solution given by setting  $x'_{\hat{f}} = x^*_{\hat{f}} - \epsilon$ , setting  $x'_f = \min(x^*_f + \epsilon, 1)$  for  $f \neq \hat{f}$ , and setting  $y'_{fv} = \min(y^*_{fv} + \epsilon, 1)$  for all  $f, v$ . This solution  $(x', y')$  remains feasible for (6.1) and costs strictly less than  $\beta$ ; therefore no such  $\hat{f}$  exists.

Having established that  $c_f \leq \beta$  for all  $f$ , we know that  $\text{support}(x^*) \subseteq \text{support}(x^0)$ . We then bound the initial  $KL$ -divergence term by

$$\begin{aligned} \text{KL}_c(x^* \parallel x^0) &= \sum_f c_f x^*_f \log\left(x^*_f \frac{c_f m'}{\beta}\right) + \sum_f c_f (x^0_f - x^*_f) \\ &\leq \beta(\log m + 1), \end{aligned}$$

where we used that  $x^*_f c_f \leq \beta$  above, that  $\langle c, x^* \rangle = \text{LP}_{\text{OPT}} \leq \beta$ , and that  $\langle c, x^0 \rangle = \beta$ .

Next we consider the second term in  $\Phi_L$ . This is the second half of the objective of the fractional relaxation (6.1), evaluated at the optimal solution; therefore it is bounded above by  $\text{LP}_{\text{OPT}} \leq \beta$ .

We turn to the last term,  $\beta \cdot \log(\sum_v \kappa_v^0 / \beta + 1/n)$ . In order to show that this is at most  $O(\beta \cdot \log n)$ , it suffices to demonstrate that  $\kappa_v^0 \leq \beta$  for all clients  $v$ . This can be seen by considering the relaxation of (6.1) to serving only the client  $v$ , which is the problem of fractionally finding the cheapest augmentation for  $v$  at time 0. Since this is a relaxation of (6.1), its optimal solution will cost at most  $\beta$ . Finally, this relaxation integrally chooses the facility  $f_v := \arg \min_f c_f + c_{fv}$ ; therefore  $\kappa_v^0 \leq \beta$ . (Since augmentation costs only decrease, indeed  $\kappa_v^t \leq \kappa_v^0 \leq \beta$  for all  $v$  and  $t$ .)

We conclude with the lower bound on  $\Phi$ . Both terms in  $\Phi_L$  are nonnegative, and so  $\Phi_C \geq -\beta \cdot \log n$  since  $\rho^t \geq 0$ .  $\square$

LEMMA 6.2. (CHANGE IN  $\Phi_L$ ) *For rounds when the event  $\Xi^t$  holds, the expected change in  $\Phi_L$  is*

$$\begin{aligned} &\mathbb{E}_{v^t, \mathcal{R}^t} [\Phi_L(t) - \Phi_L(t-1) \mid x^{t-1}, U^{t-1}, \Xi^t] \\ (6.2) \quad &\leq \mathbb{E}_{v \sim U^{t-1}} \left[ \frac{e^2 - 1}{2} \cdot \kappa_v^{t-1} \cdot \min(X^t(v), 1) - \kappa_v^{t-1} \right]. \end{aligned}$$

When the arriving  $v^t$  is preemptively connected and  $\Xi^t$  does not hold,

$$(6.3) \quad \Phi_L(t) - \Phi_L(t-1) \leq 0.$$

*Proof.* Inequality (6.3) is straightforward: when Line 5 does not execute, there is no change to  $x^t$ , and so the KL term is unchanged. At the same time, the fractional optimum term only decreases.

Our main task is to prove (6.2). We break the proof into cases. Let  $\Lambda^t$  be the event that  $X^t(v^t) < 1$ . If  $\Lambda^t$  does not hold, in Line 12 we set the vector  $x^t = x^{t-1}$ , so the change in KL term is again unchanged. This means that inequality (6.2) holds trivially, since  $(e^2 - 1)/2 > 1$ . Henceforth we focus on the case when  $\Lambda^t$  holds.

Recall that the expected change in relative entropy depends only on the arriving uncovered element  $v^t$ . Beginning with the KL term and expanding definitions, and writing  $v = v^t$  when it is clear from context,

$$\begin{aligned} &\mathbb{E}_{v^t, \mathcal{R}^t} [\text{KL}_c(x^* \parallel x^t) - \text{KL}_c(x^* \parallel x^{t-1}) \mid x^{t-1}, U^{t-1}, \Lambda^t] \\ &= \mathbb{E}_{v \sim U^{t-1}} \left[ \sum_f c_f \cdot x^*_f \cdot \log \frac{x_f^{t-1}}{x_f^t} \mid \Xi^t, \Lambda^t \right] \\ &= \mathbb{E}_{v \sim U^{t-1}} \left[ \langle c, x^* \rangle \cdot \log Z^t - \sum_{f \in \Gamma^{t-1}(v)} c_f \cdot x^*_f \cdot \log e^{\kappa_v^{t-1}/c_f} \mid \Xi^t, \Lambda^t \right] \\ (A.1) \quad &\leq \mathbb{E}_{v \sim U^{t-1}} \left[ \beta \cdot \log \left( \sum_{f \in \Gamma^{t-1}(v)} \frac{c_f}{\beta} \cdot x_f^{t-1} \cdot e^{\kappa_v^{t-1}/c_f} + \sum_{f \notin \Gamma^{t-1}(v)} \frac{c_f}{\beta} \cdot x_f^{t-1} \right) \right. \\ &\quad \left. - \sum_{f \in \Gamma^{t-1}(v)} \kappa_v^{t-1} \cdot x^*_f \right] \mid \Xi^t, \Lambda^t, \end{aligned}$$



where in the last step (A.1) we expanded the definition of  $Z^t$ , and used  $\langle c, x^* \rangle \leq \beta$ . Then we can further bound (A.1) by

$$(A.2) \quad \leq \mathbb{E}_{v \sim U^{t-1}} \left[ \beta \cdot \log \left( \sum_f \frac{c_f}{\beta} \cdot x_f^{t-1} + \frac{e^2 - 1}{2} \cdot \frac{\kappa_v^{t-1}}{\beta} \cdot X^t(v) \right) - \sum_{f \in \Gamma^{t-1}(v)} \kappa_v^{t-1} \cdot x_f^* \middle| \Xi^t, \Lambda^t \right],$$

where we use the approximation  $e^a \leq 1 + (e^2 - 1) \cdot a/2$  for  $a \in [0, 2]$  (note that  $\kappa_v^{t-1}$  is the cheapest marginal connection cost for  $v$ , so for any  $f \in \Gamma^{t-1}(v)$ , meaning that  $c_{fv} \leq \kappa_v^{t-1}/2$ , we have that  $c_f \geq \kappa_v^{t-1}/2$  and thus  $\kappa_v^{t-1}/c_f \leq 2$ ). Finally, using Invariant 6.1, along with the approximation  $\log(1 + y) \leq y$ , we bound (A.2) by

$$(A.3) \quad \begin{aligned} &\leq \mathbb{E}_{v \sim U^{t-1}} \left[ \frac{e^2 - 1}{2} \cdot \kappa_v^{t-1} \cdot X^t(v) - \sum_{f \in \Gamma^{t-1}(v)} \kappa_v^{t-1} \cdot x_f^* \middle| \Xi^t, \Lambda^t \right] \\ &\leq \mathbb{E}_{v \sim U^{t-1}} \left[ \frac{e^2 - 1}{2} \cdot \kappa_v^{t-1} \cdot \min(X^t(v), 1) - \sum_{f \in \Gamma^{t-1}(v)} \kappa_v^{t-1} \cdot x_f^* \middle| \Xi^t, \Lambda^t \right], \end{aligned}$$

where (A.3) follows by the definition of the event  $\Lambda^t$ .

We now turn to the second part of  $\Phi_L$ . The change to fractional optimum term in each round is  $-2 \cdot \sum_f c_{fv^t} \cdot y_{fv^t}^*$ . Combining this with (A.3) gives

$$\begin{aligned} &\mathbb{E}_{v \sim U^{t-1}} [\Phi_L(t) - \Phi_L(t-1) \mid \Xi^t, \Lambda^t] \\ &\leq \mathbb{E}_{v \sim U^{t-1}} \left[ \kappa_v^{t-1} \cdot \frac{e^2 - 1}{2} \cdot \min(X^t(v), 1) - \left( \sum_{f \in \Gamma^{t-1}(v)} \kappa_v^{t-1} \cdot x_f^* + 2 \cdot \sum_f c_{fv^t} \cdot y_{fv^t}^* \right) \middle| \Xi^t, \Lambda^t \right]. \end{aligned}$$

Since the fractional connection  $v$  receives from outside of  $\Gamma^{t-1}(v)$  is at cost at least  $\kappa_v^{t-1}$  by definition, we may bound this by

$$(A.4) \quad \begin{aligned} &\leq \mathbb{E}_{v \sim U^{t-1}} \left[ \kappa_v^{t-1} \cdot \frac{e^2 - 1}{2} \cdot \min(X^t(v), 1) - \left( \sum_{f \in \Gamma^{t-1}(v)} \kappa_v^{t-1} \cdot x_f^* + 2 \cdot \sum_{f \notin \Gamma^{t-1}(v)} \frac{\kappa_v^{t-1}}{2} \cdot y_{fv}^* \right) \middle| \Xi^t, \Lambda^t \right] \\ &\leq \mathbb{E}_{v \sim U^{t-1}} \left[ \kappa_v^{t-1} \cdot \frac{e^2 - 1}{2} \cdot \min(X^t(v), 1) - \kappa_v^{t-1} \middle| \Xi^t, \Lambda^t \right], \end{aligned}$$

where (A.4) follows because the fractional connection  $v$  receives from outside of  $\Gamma^{t-1}(v)$  in  $(x^*, y^*)$  is at least  $1 - \sum_{f \in \Gamma^{t-1}(v)} x_f^*$ .

We have shown the lemma statement both when  $\Lambda^t$  holds and when it does not, which completes the proof.  $\square$

LEMMA 6.3. (CHANGE IN  $\Phi_C$ ) *For all rounds  $t$  for which  $\Xi^t$  holds, the expected change in  $\Phi_C$  is*

$$(6.4) \quad \mathbb{E}_{v^t, \mathcal{R}^t} [\Phi_C(t) - \Phi_C(t-1) \mid x^{t-1}, U^{t-1}, \Xi^t] \leq -\frac{1 - e^{-1}}{4} \cdot \mathbb{E}_{u \sim U^{t-1}} [\kappa_u^{t-1} \cdot \min(X^t(v), 1)].$$

For rounds in which  $\Xi^t$  does not hold,

$$(6.5) \quad \Phi_C(t) - \Phi_C(t-1) \leq 0.$$

*Proof.* (6.5) is once again straightforward, since  $\rho^t$  is monotonically decreasing in  $t$ . We therefore focus on proving (6.4).

We start by considering the expected change to  $\Phi_C$  over the randomness of the sampling, for a fixed arriving client  $v$ . Expanding definitions,

$$\mathbb{E}_{\mathcal{R}^t} [\Phi_C(t) - \Phi_C(t-1) \mid x^{t-1}, U^{t-1}, \Xi^t, v^t = v]$$

$$\begin{aligned}
 &= \beta \cdot \mathbb{E}_{\mathcal{R}^t} \left[ \log \left( 1 - \frac{\rho^{t-1} - \rho^t}{\rho^{t-1} + \frac{\beta}{n}} \right) \middle| U^{t-1}, \Xi^t, v^t = v \right] \\
 \text{(A.5)} \quad &\leq -\beta \cdot \frac{1}{\rho^{t-1} + \frac{\beta}{n}} \cdot \mathbb{E}_{\mathcal{R}^t} [\rho^{t-1} - \rho^t \mid U^{t-1}, \Xi^t, v^t = v].
 \end{aligned}$$

Above, (A.5) follows from the approximation  $\log(1 - y) \leq -y$ . Expanding the definition of  $\rho^t$ , (A.5) is bounded by

$$\begin{aligned}
 &\leq -\frac{\beta}{\rho^{t-1} + \frac{\beta}{n}} \cdot \mathbb{E}_{\mathcal{R}^t} \left[ \sum_{u \in U^{t-1}} \frac{\kappa_u^{t-1}}{2} \cdot \mathbb{1}\{\Gamma^{t-1}(u) \cap \mathcal{R}^t \neq \emptyset\} \middle| U^{t-1}, \Xi^t, v^t = v \right] \\
 &= -\frac{\beta}{\rho^{t-1} + \frac{\beta}{n}} \sum_{u \in U^{t-1}} \frac{\kappa_u^{t-1}}{2} \cdot \mathbb{P}(\Gamma^{t-1}(u) \cap \mathcal{R}^t \neq \emptyset \mid U^{t-1}, \Xi^t, v^t = v) \\
 \text{(A.6)} \quad &\leq -\frac{\beta}{\rho^{t-1} + \frac{\beta}{n}} \sum_{u \in U^{t-1}} \frac{\kappa_u^{t-1}}{2} \cdot (1 - e^{-1}) \cdot \min \left( \frac{\kappa_v^{t-1}}{\beta} \cdot \sum_{f \in \Gamma^{t-1}(u)} x_f^{t-1}, 1 \right)
 \end{aligned}$$

$$\text{(A.7)} \quad \leq -\frac{1 - e^{-1}}{2} \cdot \kappa_v^{t-1} \cdot \frac{|U^{t-1}|}{\rho^{t-1} + \frac{\beta}{n}} \cdot \mathbb{E}_{u \sim U^{t-1}} [\kappa_u^{t-1} \cdot \min(X^t(u), 1)].$$

Step (A.6) is due to the fact that each facility  $f$  is sampled independently with probability  $\min(\kappa_v^{t-1} x_f^{t-1} / \beta, 1)$ , so the probability any given client  $u \in U^{t-1}$  gets at least one facility from  $\Gamma^{t-1}(u)$  is

$$\begin{aligned}
 1 - \prod_{f \in \Gamma^{t-1}(u)} \left( 1 - \min \left( \frac{\kappa_v^{t-1} x_f^{t-1}}{\beta}, 1 \right) \right) &\geq 1 - \exp \left\{ -\min \left( \frac{\kappa_v^{t-1}}{\beta} \sum_{f \in \Gamma^{t-1}(u)} x_f^{t-1}, 1 \right) \right\} \\
 &\stackrel{(**)}{\geq} (1 - e^{-1}) \cdot \min \left( \frac{\kappa_v^{t-1}}{\beta} \sum_{f \in \Gamma^{t-1}(u)} x_f^{t-1}, 1 \right).
 \end{aligned}$$

Above, (\*\*) follows from convexity of the exponential. Step (A.7) then follows by rewriting the sum as an expectation and using the fact that  $\kappa_v^{t-1} / \beta \leq 1$ , which is justified (and used) in the proof of Lemma 6.1.

Taking the expectation of (A.7) over  $v^t \sim U^{t-1}$ , and using the fact that  $\mathbb{E}_{v \sim U^{t-1}} [\kappa_v^{t-1}] = \rho^{t-1} / |U^{t-1}|$ , the expected change in  $\Phi_C$  becomes

$$\begin{aligned}
 &\mathbb{E}_{v^t, \mathcal{R}^t} [\Phi_C(t) - \Phi_C(t-1) \mid x^{t-1}, U^{t-1}, \Xi^t] \\
 &\leq -\frac{1 - e^{-1}}{2} \cdot \frac{\rho^{t-1}}{\rho^{t-1} + \frac{\beta}{n}} \cdot \mathbb{E}_{u \sim U^{t-1}} [\kappa_u^{t-1} \cdot \min(X^t(u), 1)], \\
 \text{(A.8)} \quad &\leq -\frac{1 - e^{-1}}{4} \cdot \mathbb{E}_{u \sim U^{t-1}} [\kappa_u^{t-1} \cdot \min(X^t(u), 1)],
 \end{aligned}$$

Where in (A.8) we finally use the fact that the event  $\Xi^t$  holds. Since this is the case, we know that  $v^t$  is not preemptively connected in round  $t$ , and so  $\rho^{t-1} \geq \kappa_{v^t}^{t-1} \geq \beta/t \geq \beta/n$ , and so  $\frac{\rho^{t-1}}{\rho^{t-1} + \frac{\beta}{n}} \geq \frac{1}{2}$ . This is the claimed bound.  $\square$

## A.2 Set Multicover

LEMMA 7.1. (BOUNDS ON  $\Phi$ ) *The initial potential is bounded as  $\Phi(0) = O(\beta \cdot \log mn)$ , and  $\Phi(t) \geq -\beta \cdot \log(n)$  for all rounds  $t$ .*

*Proof.* We start with the upper bound, and address each portion of the potential in turn. First,  $\text{KL}(x^* \parallel x^0) = \sum_j x_j^* \log \left( \frac{x_j^*}{x_j^0} \right) = \sum_j x_j^* \log \left( m \frac{x_j^*}{\beta} \right) \leq \sum_j x_j^* \log m \leq \beta \log m$ , since  $x_j^* \leq \text{LP}_{\text{OPT}} \leq \beta$  for all  $j$ . Second,  $\beta \cdot \log \left( \frac{\rho^0}{\beta} + \frac{1}{m} \right) \leq \beta \cdot \log \left( \sum_i \frac{b_i}{\beta} \right) \leq \beta \log n$ , since  $b_i \leq \text{LP}_{\text{OPT}} \leq \beta$  for all  $b_i$ .

We now justify the lower bound. The KL divergence is nonnegative, as is  $\rho^t$ ; therefore  $\Phi(t) \geq \beta \cdot \log(1/m) = -\beta \cdot \log m$ .  $\square$

LEMMA 7.2. (CHANGE IN  $\Phi_L$ ) *For rounds in which  $i^t$  arrives uncovered, the expected change in  $\Phi_L$  is*

$$(7.2) \quad \mathbb{E}_{i^t, \mathcal{R}^t} [\Phi_L(t) - \Phi_L(t-1) \mid x^{t-1}, U^{t-1}, d_{i^t}^t > 0] \leq \mathbb{E}_{i \sim U^{t-1}} [(e-1) \cdot \min(X_i^{t-1}, d_i^t) - d_i^t].$$

When  $i^t$  is covered on arrival,

$$(7.3) \quad \Phi_L(t) - \Phi_L(t-1) \leq 0.$$

*Proof.* We first show (7.3). This holds because if  $i^t$  is covered on arrival then

We break the proof of (7.2) into two cases. If  $X_i^{t-1} \geq d_i^t$ , by Line 9, the vector  $x^t$  is not updated in round  $t$ , so the change in KL divergence is 0 which means that

$$(A.9) \quad \begin{aligned} & \mathbb{E}_{i^t, \mathcal{R}^t} [\text{KL}(x^* \parallel x^t) - \text{KL}(x^* \parallel x^{t-1}) \mid x^{t-1}, U^{t-1}, X_i^{t-1} \geq d_i^t] \\ & \leq \mathbb{E}_{i \sim U^{t-1}} [(e-1) \cdot \min(X_i^{t-1}, d_i^t) - d_i^t \mid X_i^{t-1} \geq d_i^t], \end{aligned}$$

implying (7.2) trivially. Henceforth we focus on the case  $X_i^{t-1} < d_i^t$ .

Recall that the expected change in relative entropy depends only on the arriving uncovered element  $i^t$ . Expanding definitions,

$$(A.10) \quad \begin{aligned} & \mathbb{E}_{i^t, \mathcal{R}^t} [\text{KL}(x^* \parallel x^{t+1}) - \text{KL}(x^* \parallel x^t) \mid x^{t-1}, U^{t-1}, X_i^{t-1} < d_i^t] \\ & = \mathbb{E}_{i \sim U^{t-1}} \left[ \sum_j x_j^* \cdot \log \frac{x_j^{t-1}}{x_j^t} \mid X_i^{t-1} < d_i^t \right] \\ & = \mathbb{E}_{i \sim U^{t-1}} \left[ \sum_j x_j^* \cdot \log Z^t - \sum_{j \in \mathcal{T}^t} x_j^* \cdot \log e \mid X_i^{t-1} < d_i^t \right] \\ & \leq \mathbb{E}_{i \sim U^{t-1}} \left[ \beta \cdot \log Z^t - \sum_{j \in \mathcal{T}^t} x_j^* \mid X_i^{t-1} < d_i^t \right] \\ (A.11) \quad & \leq \mathbb{E}_{i \sim U^{t-1}} [\beta \cdot \log Z^t - d_i^t \mid X_i^{t-1} < d_i^t], \end{aligned}$$

where in (A.10) we used  $\langle 1, x^* \rangle \leq \beta$ . Since  $x^*$  is a feasible fractional set cover, we know that  $\sum_j a_{ij} x_j^* \geq b_i$ , and since  $x_j^* \leq 1$  this implies that  $\sum_{j \in \mathcal{T}^t} x_j^* \geq d_i^t$ , giving (A.11). Expanding  $Z^t$ , we have that

$$(A.12) \quad \begin{aligned} & = \mathbb{E}_{i \sim U^{t-1}} \left[ \beta \cdot \log \left( \frac{1}{\beta} \sum_j x_j^{t-1} + \frac{(e-1)}{\beta} \sum_{j \in \mathcal{T}^t} x_j^{t-1} \right) - d_i^t \mid X_i^{t-1} < d_i^t \right] \\ & = \mathbb{E}_{i \sim U^{t-1}} \left[ \beta \cdot \log \left( 1 + \frac{(e-1)}{\beta} X_i^{t-1} \right) - d_i^t \mid X_i^{t-1} < d_i^t \right] \end{aligned}$$

Finally the approximation  $\log(1+y) \leq y$  allows us to bound (A.12) by

$$(A.13) \quad \begin{aligned} & \leq \mathbb{E}_{i \sim U^{t-1}} [(e-1) \cdot X_i^{t-1} - d_i^t \mid X_i^{t-1} < d_i^t] \\ & = \mathbb{E}_{i \sim U^{t-1}} [(e-1) \cdot \min(X_i^{t-1}, d_i^t) - d_i^t \mid X_i^{t-1} < d_i^t]. \end{aligned}$$

The lemma statement follows by combining (A.9) and (A.13) using the law of total expectation.  $\square$

LEMMA 7.3. (CHANGE IN  $\Phi_C$ ) *In every round for which  $i^t$  is uncovered on arrival, the expected change in  $\Phi_C$  is*

$$(7.4) \quad \mathbb{E}_{i^t, \mathcal{R}^t} [\Phi_C(t) - \Phi_C(t-1) \mid x^{t-1}, U^{t-1}, d_{i^t}^t > 0] \leq -\frac{\gamma}{2} \cdot \mathbb{E}_{i \sim U^{t-1}} [\min(X_i^{t-1}, d_i^t)],$$

where  $\gamma$  is a fixed constant. On rounds in which  $i^t$  arrives covered,

$$(7.5) \quad \Phi_C(t) - \Phi_C(t-1) \leq 0.$$

We will require an additional fact, the proof of which appears in [GKL21, Appendix A]:

FACT A.1. *Given probabilities  $p_j$  and coefficients  $b_j \in [0, 1]$ , let  $W := \sum_j b_j \text{Ber}(p_j)$  be the sum of independent weighted Bernoulli random variables. Let  $\Delta \geq (e-1)^{-1}$  be some constant. Then*

$$\mathbb{E}[\min(W, \Delta)] \geq \gamma \cdot \min(\mathbb{E}[W], \Delta),$$

for a fixed constant  $\gamma$  independent of the  $p_j$  and  $b_j$ .

*Proof of Lemma 7.3.* When  $i^t$  is covered on arrival (7.5) holds trivially, since in this case  $d_{i^t}^t = 0$  and so no sets  $j$  are bought and  $\rho^t = \rho^{t-1}$ . We therefore focus on proving (7.4) in the case when  $i^t$  is uncovered on arrival.

Conditioned on  $i^t = i$ , the expected change in  $\log \rho^t$  depends only on  $\mathcal{R}^t$ :

$$(A.14) \quad \begin{aligned} & \mathbb{E}_{i^t, \mathcal{R}^t} [\Phi_C(t) - \Phi_C(t-1) \mid x^{t-1}, U^{t-1}, i^t = i] \\ &= \beta \cdot \mathbb{E}_{\mathcal{R}^t} \left[ \log \left( 1 - \frac{\rho^{t-1} - \rho^t}{\rho^{t-1} + \frac{\beta}{m}} \right) \mid U^{t-1}, i^t = i \right] \\ &\leq -\beta \cdot \frac{1}{\rho^{t-1} + \frac{\beta}{m}} \mathbb{E}_{\mathcal{R}^t} [\rho^{t-1} - \rho^t \mid U^{t-1}, i^t = i]. \end{aligned}$$

Above, (A.14) follows from the approximation  $\log(1-y) \leq -y$ . Expanding definitions again, we further bound (A.14) by

$$(A.15) \quad \begin{aligned} &= -\beta \cdot \frac{1}{\rho^{t-1} + \frac{\beta}{m}} \mathbb{E}_{\mathcal{R}^t} \left[ \sum_{i' \in U^{t-1}} (d_{i'}^{t-1} - d_{i'}^t) \right] \\ &= -\beta \cdot \frac{1}{\rho^{t-1} + \frac{\beta}{m}} \sum_{i' \in U^{t-1}} \mathbb{E}_{\mathcal{R}^t} [d_{i'}^{t-1} - d_{i'}^t] \end{aligned}$$

$$(A.16) \quad \leq -\beta \cdot \frac{1}{\rho^{t-1} + \frac{\beta}{m}} \sum_{i' \in U^{t-1}} \mathbb{E}_{\mathcal{R}^t} \left[ \min \left( \sum_{j \in \mathcal{T}^t} \text{Ber}(x_j^{t-1} \cdot d_i^{t-1} / \beta), d_{i'}^t \right) \right].$$

Here (A.16) follows from the preceding line by the definition of the random sampling performed in Line 8 (with inequality because the algorithm buys other coordinates also). This Bernoulli probability is well-defined because  $d_{i'}^{t-1} \leq \text{LP}_{\text{OPT}} \leq \beta$ , and since Line 10 guarantees that  $x_j^{t-1} \leq 1$  for all  $j$  in  $\mathcal{T}^t$ . The expectation of this sum of Bernoullis is  $\sum_{j \in \mathcal{T}^t} x_j^{t-1} d_i^{t-1} / \beta = X_i^{t-1} d_i^{t-1} / \beta$ . Since  $d_i^{t-1} \geq 1$ , we may apply Lemma A.1 to obtain

$$(A.17) \quad \begin{aligned} &\leq -\frac{1}{\rho^{t-1} + \frac{\beta}{m}} \sum_{i' \in U^{t-1}} \gamma \cdot \min \left( \frac{d_i^{t-1}}{\beta} \cdot X_{i'}^{t-1}, d_{i'}^t \right) \\ &\leq -\gamma \cdot d_i^{t-1} \cdot \frac{1}{\rho^{t-1} + \frac{\beta}{m}} \sum_{i' \in U^{t-1}} \min(X_{i'}^{t-1}, d_{i'}^t) \\ &= -\gamma \cdot d_i^{t-1} \cdot \frac{|U^{t-1}|}{\rho^{t-1} + \frac{\beta}{m}} \cdot \mathbb{E}_{i' \sim U^{t-1}} [\min(X_{i'}^{t-1}, d_{i'}^t)] \end{aligned}$$

$$(A.18) \quad \leq -\gamma \cdot d_i^{t-1} \cdot \frac{|U^{t-1}|}{2 \cdot \rho^{t-1}} \cdot \mathbb{E}_{i' \sim U^{t-1}} [\min(X_{i'}^{t-1}, d_{i'}^{t-1})],$$

where (A.18) follows from the observation that  $i^t$  arrives uncovered, so  $\rho^{t-1} \geq d_i^t \geq 1$  while  $\beta/m \leq 1$ ; therefore  $\frac{1}{\rho^{t-1} + \beta/m} \leq \frac{1}{2 \cdot \rho^{t-1}}$ . Taking the expectation of (A.18) over  $i \sim U^{t-1}$ , and using the fact that  $\mathbb{E}_{i \sim U^{t-1}} [d_i^{t-1}] = \rho^{t-1}/|U^{t-1}|$ , the expected change in  $\log \rho^t$  becomes

$$\mathbb{E}_{i^t, \mathcal{R}^t} [\Phi_C(t) - \Phi_C(t-1) \mid x^{t-1}, U^{t-1}] \leq -\frac{\gamma}{2} \cdot \mathbb{E}_{i \sim U^{t-1}} [\min(X_i^{t-1}, d_i^t)],$$

as desired.  $\square$

## B Lower Bounds for Relaxed Models

**B.1 Adversarial Corruptions** It is tempting to try to extend the with-a-sample model to the case where the samples are noisy. In this section, we study one natural model, and show that, sadly, no randomized algorithm can achieve a competitive ratio of  $o(\log m \log n)$  in polynomial time, unless  $\mathsf{P} = \mathsf{NP}$ .

We begin by describing the model, which we call  $(\alpha, \delta)$ -NOISYSAMPLESETCOVER. The adversary begins by committing to an online set cover input sequence of length  $n$  from a set system with  $N$  elements and  $m$  sets. Next a uniformly random subset  $S_1$  of the input is sampled, with  $|S_1| = \alpha \cdot n$ . The adversary then chooses a set  $S_2$  of size at most  $\delta \cdot n$ , independently of  $S_1$ . The algorithm is given the set of constraints  $S = S_1 \cup S_2$  in advance. Finally, the online sequence begins. In this model  $\alpha$  is the proportion of the online input sampled (as before),  $\delta$  represents the rate of corruption that the adversary can introduce to the sample solely by adding additional constraints, and  $S_2$  the adversary's choice of additions. (A very similar construction demonstrates lower bounds in an alternative model where the adversary may instead omit already-sampled constraints and the algorithm is given the set  $S = S_1 \setminus S_2$ .)

We now construct our hard instance for  $(\alpha, \delta)$ -NOISYSAMPLESETCOVER when  $\alpha = \Omega(1)$  and  $\delta = n^{-1+\Omega(1)}$ . We will use as a sub-instance the construction from [Kor04] which shows an  $\Omega(\log m \log n)$  lower bound for the original online set cover problem, which holds in the setting where the universe of elements is known (though the arriving subset and order are not). Let  $(\mathcal{U}_{hard}, \mathcal{S}_{hard})$  be the underlying set system in the [Kor04] instance. Let  $\sigma_{hard}$  be the sequence of elements given to the algorithm, and let the unordered set of these elements be called  $U_{hard} \subseteq \mathcal{U}_{hard}$ . In the construction of [Kor04],  $U_{hard}$  is a random variable such that  $|U_{hard}| = \Theta(\sqrt{|\mathcal{U}_{hard}|})$ .

Our construction is the following. The set system consists of a set  $S_0$  containing  $N - \delta \cdot n$  elements, together with a copy of the set system  $(\mathcal{U}_{hard}, \mathcal{S}_{hard})$  with parameters  $n' := |U_{hard}| = \delta \cdot n$  and  $m' := |\mathcal{S}_{hard}| = \text{poly}(n')$ . The adversary commits to the online sequence which reveals all the elements of  $S_0$  in arbitrary order, and then the elements of  $U_{hard} \subseteq \mathcal{U}_{hard}$  in the order  $\sigma_{hard}$ . This sequence is of length  $n = \Theta(N)$  (provided  $\delta$  is bounded away from 1). Some sample  $S_1$  of this input sequence is drawn. The adversary then picks  $S_2$ , which is of size  $\delta \cdot n$ , to be all the elements of  $\mathcal{U}_{hard}$ . Hence no matter what the realization of the sample  $S_1$  is, the algorithm has no information about the identity of  $U_{hard} \subseteq \mathcal{U}_{hard}$  and must cover the hard online set cover sequence  $\sigma_{hard}$  of length  $\Theta(\sqrt{n'}) = n^{\Omega(1)}$  with no useful advice. Since any polynomial-time randomized online set cover algorithm has competitive ratio  $\Omega(\log n' \log m')$  on  $\sigma_{hard}$ , no algorithm can achieve competitive ratio  $o(\log n \log m)$  for  $(\alpha, \delta)$ -NOISYSAMPLESETCOVER when  $\alpha = \Omega(1)$  and  $\delta = n^{-1+\Omega(1)}$ .

**B.2 Relaxed Random Order** Another interesting question is whether [GKL21] can be made to work when the input ordering is not fully uniformly random, but only nearly so.

We show that the entropy of the arrival order distribution is *not* a good parametrization of the distance to random order, in that there exist instances and distributions over arrival orders with nearly full entropy,  $(1 - \epsilon)n \log n \approx \log(n!)$ , but for which any online algorithm has competitive ratio  $\Omega(\log(\epsilon m) \log(\epsilon n))$ .

One simple such instance is the following. There are  $(1 - \epsilon)n$  dummy elements presented in uniformly random order, followed by a hard online set cover sequence  $\sigma_{hard}$  of length  $\epsilon n$ . The permutation distribution has the desired near maximal entropy, but no randomized polynomial time algorithm has competitive ratio  $o(\log(\epsilon m) \log(\epsilon n))$  unless  $\mathsf{P} = \mathsf{NP}$ , by the lower bound of [Kor04].

## C Error in [DEH<sup>+</sup>18b]

In [DEH<sup>+</sup>18b] the authors claim an  $O(\log n)$ -competitive algorithm for prophet set cover via a reduction from prophet set cover to known i.i.d. set cover, which is the special case when all distributions are identical. (They also claim an  $O(1)$ -competitive algorithm for prophet METRICFACILITYLOCATION via the same reduction.) Their proof of this reduction, which appears in [EB17, Theorem 9.10] relies on the following claim.

Let  $D^1, \dots, D^n$  be a sequence of distributions over elements of  $\mathcal{U}$ , and let  $D^* = \frac{1}{n} \sum_{i=1}^n D^i$  be the average distribution. Let  $\text{OPT}_{pht}$  be the expected size of the optimal set cover for  $U$  when  $U$  is formed by drawing one element from each of  $D^1, \dots, D^n$ . Let  $\text{OPT}_{iid}$  be the size of the optimal set cover when  $U$  is formed by drawing  $n$  times from  $D^*$ . The claim is that  $\text{OPT}_{iid} \leq \text{OPT}_{pht}$ , and we demonstrate that this does not hold in general.

Consider the instance with universe  $(i, j)$  for  $i \in [2]$  and  $j \in [2]$  and the set system

$$\mathcal{S} = \{(1, 1), (2, 1)\}, \{(1, 1), (2, 2)\}, \{(1, 2), (2, 1)\}, \{(1, 2), (2, 2)\}.$$

Let the prophet distributions  $D^1$  and  $D^2$ , be the uniform distributions over  $(1, 1), (1, 2)$  and  $(2, 1), (2, 2)$ . Let  $D^* = \frac{1}{2}(D^1 + D^2)$  be their average. Then  $\text{OPT}_{pht} = 1$ , while  $\mathbb{E}[\text{OPT}_{iid}] = 5/4$ . More generally, let the universe be  $(i, j)$  for  $i \in [n]$  and  $j \in [\log n]$  and the collection of sets be all sequences  $\mathcal{S} = [\log n]^{[n]}$ . Then the  $n$  prophet distributions  $D^i$  are each the uniform distribution over  $\{(i, 1), (i, 2), \dots, (i, \log n)\}$ , and the average distribution is  $D^* = \frac{1}{n} \sum_i D^i$  as before. A balls-and-bins argument shows that

$$\text{OPT}_{pht} = 1, \quad \mathbb{E}[\text{OPT}_{iid}] = \Omega\left(\frac{\log n}{\log \log n}\right),$$

which demonstrates that the claimed inequality is in the worst case violated by at least this multiplicative factor.

## References

- [AFGS22] C.J. Argue, Alan Frieze, Anupam Gupta, and Christopher Seiler. Learning from a sample in online algorithms. In *NeurIPS*, Dec 2022.
- [AKW14] Pablo Daniel Azar, Robert Kleinberg, and S. Matthew Weinberg. Prophet inequalities with limited information. In Chandra Chekuri, editor, *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5-7, 2014*, pages 1358–1377. SIAM, 2014.
- [BC22] Archit Bubna and Ashish Chiplunkar. Prophet inequality: Order selection beats random order. *CoRR*, abs/2211.04145, 2022.
- [BL11] John R Birge and Francois Louveaux. *Introduction to stochastic programming*. Springer Science & Business Media, 2011.
- [BN09] Niv Buchbinder and Joseph Naor. Online primal-dual algorithms for covering and packing. *Math. Oper. Res.*, 34(2):270–286, 2009.
- [CCP05] Moses Charikar, Chandra Chekuri, and Martin Pál. Sampling bounds for stochastic optimization. In Chandra Chekuri, Klaus Jansen, José D. P. Rolim, and Luca Trevisan, editors, *Approximation, Randomization and Combinatorial Optimization, Algorithms and Techniques, 8th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems, APPROX 2005 and 9th International Workshop on Randomization and Computation, RANDOM 2005, Berkeley, CA, USA, August 22-24, 2005, Proceedings*, volume 3624 of *Lecture Notes in Computer Science*, pages 257–269. Springer, 2005.
- [CCZ14] Michele Conforti, Gérard Cornuéjols, and Giacomo Zambelli. *Integer Programming*. Springer International Publishing, 2014.
- [Chv79] Vasek Chvatal. A greedy heuristic for the set-covering problem. *Mathematics of operations research*, 4(3):233–235, 1979.
- [DEH<sup>+</sup>18a] Sina Dehghani, Soheil Ehsani, MohammadTaghi Hajiaghayi, Vahid Liaghat, and Saeed Seddighin. Greedy algorithms for online survivable network design. In Ioannis Chatzigiannakis, Christos Kaklamanis, Daniel Marx, and Donald Sannella, editors, *45th International Colloquium on Automata, Languages, and Programming, ICALP 2018, July 9-13, 2018, Prague, Czech Republic*, volume 107 of *LIPICs*, pages 152:1–152:14. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018.
- [DEH<sup>+</sup>18b] Sina Dehghani, Soheil Ehsani, MohammadTaghi Hajiaghayi, Vahid Liaghat, and Saeed Seddighin. Greedy algorithms for online survivable network design. In *45th International Colloquium on Automata, Languages, and Programming (ICALP 2018)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018.

- [DS14] Irit Dinur and David Steurer. Analytical Approach to Parallel Repetition. In *Proceedings of the Forty-sixth Annual ACM Symposium on Theory of Computing*, STOC '14, pages 624–633, New York, NY, USA, 2014. ACM.
- [EB17] Soheil Ehsani Banafati. *Online Decision Making via Prophet Setting*. PhD thesis, University of Maryland, 2017.
- [F<sup>+</sup>89] Thomas S Ferguson et al. Who solved the secretary problem? *Statistical science*, 4(3):282–289, 1989.
- [Fei98] Uriel Feige. A threshold of  $\ln n$  for approximating set cover. *J. ACM*, 45(4):634–652, 1998.
- [GGL<sup>+</sup>13] Fabrizio Grandoni, Anupam Gupta, Stefano Leonardi, Pauli Miettinen, Piotr Sankowski, and Mohit Singh. Set covering with our eyes closed. *SIAM J. Comput.*, 42(3):808–830, 2013.
- [GGLS08] Naveen Garg, Anupam Gupta, Stefano Leonardi, and Piotr Sankowski. Stochastic analyses for online combinatorial optimization problems. In Shang-Hua Teng, editor, *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2008, San Francisco, California, USA, January 20-22, 2008*, pages 942–951. SIAM, 2008.
- [GKL21] Anupam Gupta, Gregory Kehne, and Roie Levin. Random order online set cover is as easy as offline. In *62nd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2021, Denver, CO, USA, February 7-10, 2022*, pages 1253–1264. IEEE, 2021.
- [GPRS04] Anupam Gupta, Martin Pál, R. Ravi, and Amitabh Sinha. Boosted sampling: approximation algorithms for stochastic optimization. In László Babai, editor, *Proceedings of the 36th Annual ACM Symposium on Theory of Computing, Chicago, IL, USA, June 13-16, 2004*, pages 417–426. ACM, 2004.
- [GPRS05] Anupam Gupta, Martin Pál, R. Ravi, and Amitabh Sinha. What about wednesday? approximation algorithms for multistage stochastic optimization. In Chandra Chekuri, Klaus Jansen, José D. P. Rolim, and Luca Trevisan, editors, *Approximation, Randomization and Combinatorial Optimization, Algorithms and Techniques, 8th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems, APPROX 2005 and 9th International Workshop on Randomization and Computation, RANDOM 2005, Berkeley, CA, USA, August 22-24, 2005, Proceedings*, volume 3624 of *Lecture Notes in Computer Science*, pages 86–98. Springer, 2005.
- [GRS07] Anupam Gupta, R. Ravi, and Amitabh Sinha. LP rounding approximation algorithms for stochastic network design. *Math. Oper. Res.*, 32(2):345–364, 2007.
- [HK92] Theodore P Hill and Robert P Kertz. A survey of prophet inequalities in optimal stopping theory. *Contemp. Math*, 125:191–207, 1992.
- [JLN<sup>+</sup>05] Lujun Jia, Guolong Lin, Guevara Noubir, Rajmohan Rajaraman, and Ravi Sundaram. Universal approximations for tsp, steiner tree, and set cover. In Harold N. Gabow and Ronald Fagin, editors, *Proceedings of the 37th Annual ACM Symposium on Theory of Computing, Baltimore, MD, USA, May 22-24, 2005*, pages 386–395. ACM, 2005.
- [Joh74] David S Johnson. Approximation algorithms for combinatorial problems. *Journal of computer and system sciences*, 9(3):256–278, 1974.
- [KNR20] Haim Kaplan, David Naori, and Danny Raz. Competitive analysis with a sample and the secretary problem. In Shuchi Chawla, editor, *Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms, SODA 2020, Salt Lake City, UT, USA, January 5-8, 2020*, pages 2082–2095. SIAM, 2020.
- [KNR22] Haim Kaplan, David Naori, and Danny Raz. Online weighted matching with a sample. In Joseph (Seffi) Naor and Niv Buchbinder, editors, *Proceedings of the 2022 ACM-SIAM Symposium on Discrete Algorithms, SODA 2022, Virtual Conference / Alexandria, VA, USA, January 9 - 12, 2022*, pages 1247–1272. SIAM, 2022.
- [KNR23] Haim Kaplan, David Naori, and Danny Raz. Almost tight bounds for online facility location in the random-order model. In Nikhil Bansal and Viswanath Nagarajan, editors, *Proceedings of the 2023 ACM-SIAM Symposium on Discrete Algorithms, SODA 2023, Florence, Italy, January 22-25, 2023*, pages 1523–1544. SIAM, 2023.
- [Kor04] Simon Korman. On the use of randomization in the online set cover problem. *Master's thesis, Weizmann Institute of Science, Rehovot, Israel*, 2004.
- [KS78] Ulrich Krengel and Louis Sucheston. On semiamarts, amarts, and processes with finite value. *Probability on Banach spaces*, 4:197–266, 1978.
- [KT84] Antoon WJ Kolen and Arie Tamir. *Covering problems*. Econometric Institute, 1984.
- [LLP<sup>+</sup>21] Allen Liu, Renato Paes Leme, Martin Pál, Jon Schneider, and Balasubramanian Sivan. Variable decomposition for prophet inequalities and optimal ordering. In Péter Biró, Shuchi Chawla, and Federico Echenique, editors, *EC '21: The 22nd ACM Conference on Economics and Computation, Budapest, Hungary, July 18-23, 2021*, page 692. ACM, 2021.
- [Lov75] László Lovász. On the ratio of optimal integral and fractional covers. *Discrete mathematics*, 13(4):383–390, 1975.
- [Mey01] Adam Meyerson. Online facility location. In *Proceedings 42nd IEEE Symposium on Foundations of Computer Science*, pages 426–431. IEEE, 2001.
- [PT22] Bo Peng and Zhihao Gavin Tang. Order selection prophet inequality: From threshold optimization to arrival time design. In *63rd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2022, Denver, CO, USA, October 31 - November 3, 2022*, pages 171–178. IEEE, 2022.
- [RWW20] Aviad Rubinfeld, Jack Z. Wang, and S. Matthew Weinberg. Optimal single-choice prophet inequalities from samples. In Thomas Vidick, editor, *11th Innovations in Theoretical Computer Science Conference, ITCS 2020, January*

12-14, 2020, Seattle, Washington, USA, volume 151 of *LIPICs*, pages 60:1–60:10. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020.

- [Sed] Saeed Seddighin. personal communication.
- [SS06a] David B. Shmoys and Chaitanya Swamy. An approximation scheme for stochastic linear programming and its application to stochastic integer programs. *J. ACM*, 53(6):978–1012, 2006.
- [SS06b] Chaitanya Swamy and David B. Shmoys. Approximation algorithms for 2-stage stochastic optimization problems. In S. Arun-Kumar and Naveen Garg, editors, *FSTTCS 2006: Foundations of Software Technology and Theoretical Computer Science, 26th International Conference, Kolkata, India, December 13-15, 2006, Proceedings*, volume 4337 of *Lecture Notes in Computer Science*, pages 5–19. Springer, 2006.
- [Vyg05] Jens Vygen. *Approximation algorithms facility location problems*. Forschungsinstitut für Diskrete Mathematik, Rheinische Friedrich-Wilhelms ..., 2005.
- [WS11] David P Williamson and David B Shmoys. *The design of approximation algorithms*. Cambridge university press, 2011.