Nearly Minimax Optimal Submodular Maximization with Bandit Feedback

Artin Tajdini, Lalit Jain, Kevin Jamieson
University of Washington, Seattle, WA
{artin, jamieson}@cs.washington.edu, lalitj@uw.edu

Abstract

We consider maximizing an unknown monotonic, submodular set function $f : 2^{[n]} \to [0,1]$ with cardinality constraint under stochastic bandit feedback. At each time $t = 1, \dots, T$ the learner chooses a set $S_t \subset [n]$ with $|S_t| \leq k$ and receives reward $f(S_t) + \eta_t$ where η_t is mean-zero sub-Gaussian noise. The objective is to minimize the learner's regret with respect to an approximation of the maximum $f(S_*)$ with $|S_*| = k$, obtained through robust greedy maximization of f. To date, the best regret bound in the literature scales as $kn^{1/3}T^{2/3}$. And by trivially treating every set as a unique arm one deduces that $\sqrt{\binom{n}{k}T}$ is also achievable using standard multi-armed bandit algorithms. In this work, we establish the first minimax lower bound for this setting that scales like $\tilde{\Omega}(\min_{L \leq k}(L^{1/3}n^{1/3}T^{2/3} + \sqrt{\binom{n}{k-L}T}))$. For a slightly restricted algorithm class, we prove a stronger regret lower bound of $\tilde{\Omega}(\min_{L \leq k}(Ln^{1/3}T^{2/3} + \sqrt{\binom{n}{k-L}T}))$. Moreover, we propose an algorithm Sub-UCB that achieves regret $\tilde{\mathcal{O}}(\min_{L \leq k}(Ln^{1/3}T^{2/3} + \sqrt{\binom{n}{k-L}T}))$ capable of matching the lower bound on regret for the restricted class up to logarithmic factors.

1 INTRODUCTION

Optimizing over sets of n ground items given noisy feedback is a common problem. For example, when a patient comes into the hospital with sepsis (bacterial infection of the blood), it is common for a cocktail of $1 < k \leq n$ antibiotics to be prescribed. This can be attractive for reasons including 1) the set could be as effective (or more) than a single drug alone, but each unit of the cocktail could be administered at a far lower dosage to avoid toxicity, or 2) could be more robust to resistance by blocking a number of different pathways that would have to be overcome simultaneously, or 3) could cover a larger set of pathogens present in the population. In this setting the prescriber wants to balance exploration with exploitation over different subsets to maximize the number of patients that survive. As a second example, we consider factorial optimization of web-layouts: you have n pieces of content and k locations on the webpage to place them–how do you choose subsets to maximize metrics like click-through rate or engagement?

Given there are $\approx n^k$ ways to choose k items amongst a set of n, this optimization problem is daunting. It is further complicated by the fact that for any set S_t that we evaluate at time t, we only get to observe a noisy realization of f, namely $y_t = f(S_t) + \eta_t$ where η_t is mean-zero, sub-Gaussian noise. In the antibiotics case, this could be a Bernoulli indicating whether the patient recovered or not, and in the web-layout case this could be a Bernoulli indicating a click or a (clipped) real number to represent the engagement time on the website. To make this problem more tractable, practitioners make structural assumptions about f. A common assumption is to assume that higher-order interaction terms are negligible Hill et al. (2017); Chen et al. (2021). For example, assuming

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

only interactions up to the second degree would mean that there exist parameters $\theta^{(0)} \in \mathbb{R}$, $\theta^{(1)} \in \mathbb{R}^n$, and $\theta^{(2)} \in \mathbb{R}^{\binom{n}{2}}$ such that

$$f(S) = \theta^{(0)} + \sum_{i \in S} \theta^{(1)}_i + \sum_{i,j \in S, i \neq j} \theta^{(2)}_{i,j}. \tag{1}$$

However, this model can be very restrictive and even if true, the number of unknowns scales like n^2 which could still be intractably large.

An alternative strategy is to remain within a non-parametric class, but reduce our ambitions to measuring performance relative to a different benchmark which is easier to optimize. We say a set function $f : 2^{[n]} \to \mathbb{R}$ is *increasing and submodular* if for all $A \subset B \subset [n]$ we have $f(A) \leq f(B)$ and

$$f(A \cup B) + f(A \cap B) \leq f(A) + f(B). \tag{2}$$

Such a condition limits how quickly f can grow and captures some notion of diminishing returns. Diminishing returns is reasonable in both the antibiotics and webpage optimization examples. It is instructive to note that a sufficient condition for the parametric form of (1) to be submodular is for $\max_{i,j} \theta^{(2)}_{i,j} \leq 0$. But in general, f still has $\approx n^k$ degrees of freedom even if it is monotonic and submodular. And it is known that for an unknown f, identifying $S^* := \arg\max_{S \subset [n]:|S|=k} f(S)$ may require evaluating f as many as n^k times.

The power of submodularity is made apparent through the famous result of Nemhauser and Wolsey (1978) which showed that the *greedy algorithm* which grows a set one item at a time by adding the item with the highest marginal gain returns a solution that is within a $(1-e^{-1})$-multiplicative factor of the optimal solution. That is, if we begin with $S^f_{gr} = \emptyset$ and set $S^f_{gr} \leftarrow \arg\max_{i \in [n] \setminus S^f_{gr}} f(S^f_{gr} \cup \{i\})$ until $|S^f_{gr}| = k$, then $f(S^f_{gr}) \geq (1 - 1/e) f(S^f_*)$ where $S^f_* := \arg\max_{S \in [n]:|S| \leq k} f(S)$ if f is increasing and submodular. This result is complemented by Feige (1998) which shows achieving any $(1 - e^{-1} + \epsilon)$-approximation is NP-Hard. Under additional assumptions like curvature, this guarantee can be strengthened.

Due to the centrality of the greedily constructed set to the optimization of a submodular function, it is natural to define a performance measure relative to the greedily constructed set. However, as discussed at length in the next section, because we only observe noisy observations of the underlying function, recovering the set constructed greedily from noiseless evaluations is too much to hope for. Consequently, there is a more natural notion of regret against a noisy greedy solution, denoted $R_{\textbf{gr}}$, that actually appears in the proofs of all upper bounds found in the literature for this setting (see the next section for a definition).

For this notion of regret, previous works have demonstrated that a regret bound of $R_{\textbf{gr}} = O(\text{poly}(k) n^{1/3} T^{2/3})$ is achievable (Nie et al. (2022), Streeter and Golovin (2007)). This $T^{2/3}$ rate is unusual in multi-armed bandits, where frequently we expect a regret bound to scale as $T^{1/2}$. On the other hand, by treating each k-subset as a separate arm, one can easily adapt existing algorithms to achieve a regret bound of $\sqrt{\binom{n}{k} T}$. This leads to the following question:

> *Does there exist an algorithm that obtains $\sqrt{n^r T}$ regret for $r = o(k)$ on every instance? And if not, what is the optimal dependence on k and n for a bound scaling like $T^{2/3}$?*

To address these questions, we prove a minimax lower bound and complement the result with an algorithm achieving a matching upper bound. To be precise, the contributions of this paper include:

- A minimax lower bound demonstrating that $R_{\textbf{gr}} = \tilde{\Omega}\Big(\min_{0 \leq L \leq k}(L^{1/3} n^{1/3} T^{2/3} + \sqrt{\binom{n}{k-L} T)}\Big)$. In words, for small T, a $T^{2/3}$ regret bound is inevitable, for large T the $\sqrt{\binom{n}{k} T}$ bound is optimal, with an interpolating regret bound for in between.
  - For slightly restricted class of algorithms with non-adaptive greedy error threshold, we have the improved $R_{\textbf{gr}} = \tilde{\Omega}(\Big(\min_{0 \leq L \leq k}(L n^{1/3} T^{2/3} + \sqrt{\binom{n}{k-L} T)}\Big)$.

- We propose an algorithm that for any increasing, submodular f, we have $R_{\mathbf{gr}} = \tilde{\mathcal{O}} \min_{0 \le L \le k}(Ln^{1/3}T^{2/3} + \sqrt{\binom{n}{k-L}T})$. As this matches our lower bound, we conclude that this is the first provably tight algorithm for optimizing increasing, submodular functions with bandit feedback. Existing algorithms construct a set by greedily adding k items. Our main insight is that it is actually optimal to build up a set up to a size i^* and then for the remaining stages play sets of size k that include the initial set of size i^*. Our choice of i^* is directly motivated by our lower bound.

In what remains, we will formally define the problem, discuss the related work, and then move on to the statement of the main theoretical results. Experiments and conclusions follow.

1.1 Problem Statement

Let $[n] = \{1, \ldots, n\}$ denote the set of base arms, T be the time horizon, and k be a given cardinality constraint. At time t, the agent selects a set $S_t \subset [n]$ where $|S_t| \le k$, and observes reward $f(S_t) + \eta_t$ where η_t is i.i.d. mean-zero 1-sub-Gaussian noise, and $f : 2^{[n]} \to [0, 1]$ is an unknown monotone non-decreasing submodular function defined for all sets of cardinality at most k.

Ideally, our goal would be to minimize the regret relative to pulling the best set $S^* := \arg\max_{|S| \le k} f(S)$ at each time. In general, even if we had the ability to evaluate the true function $f(\cdot)$ (i.e. without noise), maximizing a submodular function with a cardinality constraint is NP-hard. However, greedy algorithms which sequentially add points, i.e. $S^{(i+1)} = \arg\max_{a \notin S^{(i)}} f(S^{(i)} \cup a), 1 \le i \le k$ guarantee that $f(S^{(k)}) \ge \alpha f(S^\star)$ with $\alpha \ge 1 - 1/e$ in worst-case. Unfortunately, since we do not know $f(\cdot)$ and instead only have access to noisy observations, running the greedy algorithm on any estimate $\hat{f}(\cdot)$ may not necessarily guarantee an $\alpha = (1 - 1/e)$-approximation to $f(S^*)$[1].

Consequently, a natural notion to address noisy observations is an $\boldsymbol{\epsilon}$-approximate greedy set for $\boldsymbol{\epsilon} \in [0, 1]^k$. We define the following collection of sets of size k

$$\mathcal{S}^{k,\boldsymbol{\epsilon}} = \{S = S^{(k)} \supset \cdots \supset S^{(1)}, |S^{(i)}| = i, \\ \max_{a \notin S^{(i)}} f(S^{(i)} \cup \{a\}) - f(S^{(i+1)}) \le \epsilon_i\}.$$

Intuitively, any $S \in \mathcal{S}^{k,\boldsymbol{\epsilon}}$ can be thought of as being constructed from a process that adds an element at stage i which is ϵ_i-optimal compared to the Greedy algorithm run on f. Such a set naturally arises as the output of the Greedy algorithm run on an approximation $\hat{f}$. This set enjoys the following guarantee.

Lemma 1.1. *(Theorem 6 in Streeter and Golovin (2007)) For any $\boldsymbol{\epsilon} \ge \mathbf{0} \in \mathbb{R}^k$, and $S^{k,\boldsymbol{\epsilon}}_{\mathbf{gr}} \in \mathcal{S}^{k,\boldsymbol{\epsilon}}$, we have*

$$f(S^{k,\boldsymbol{\epsilon}}_{\mathbf{gr}}) + \mathbf{1}^T\boldsymbol{\epsilon} \ge (1 - e^{-1})f(S^*).$$

Lemma 1.1 is a noise-robust analogous result to the approximation ratio of the perfect greedy algorithm of Nemhauser and Wolsey (1978) that says $f(S^{k,0}_{\mathbf{gr}}) \ge (1 - e^{-1})f(S^*)$. Note that $|\mathcal{S}^{k,\boldsymbol{\epsilon}}|$ is non-decreasing in ϵ_i for all $i \in [k]$, so identifying a set in $\mathcal{S}^{k,\boldsymbol{\epsilon}}$ is in some sense easier for a larger $\mathbf{1}^T\boldsymbol{\epsilon}$. Thus, to define an appropriate definition of regret, the measure must balance the facts that comparing with the noiseless greedy approximation in $\mathcal{S}^{k,\mathbf{0}}$ may be impossible, but should account for identifying a set in $\mathcal{S}^{k,\boldsymbol{\epsilon}}$ is strictly easier for larger $\mathbf{1}^T\boldsymbol{\epsilon}$. Inspired by the above lemma we define *robust greedy regret*

$$R_{\mathbf{gr}} := \min_{\boldsymbol{\epsilon} \ge \mathbf{0}, S^{k,\boldsymbol{\epsilon}}_{\mathbf{gr}} \in \mathcal{S}^{k,\boldsymbol{\epsilon}}} R(S^{k,\boldsymbol{\epsilon}}_{\mathbf{gr}}) + T\mathbf{1}^T\boldsymbol{\epsilon} \tag{3}$$

where

$$R(S) := \sum_{t=1}^{T} f(S) - f(S_t).$$

[1] The gap between maximum gain and rest of the elements in the greedy path for lower cardinalities can be arbitrary small, making them indistinguishable with T queries. Therefore, only making queries to sets of size k would give any information on the greedy solution.

Function Assumptions	Stochastic	Regret	Upper Bound	Lower Bound
Submodular+monotone	✓	$R_{\textbf{gr}}$	$kn^{1/3}T^{2/3}$ Pedramfar and Aggarwal (2023)	$\boldsymbol{\min_L(L^{1/3}n^{1/3}T^{2/3} + \sqrt{\binom{n}{k-L}T})}$ **(This work)**
Submodular+monotone	×	$R_{\textbf{gr}}$	$kn^{1/3}T^{2/3}$ Streeter and Golovin (2007)	$\boldsymbol{\min_L(L^{1/3}n^{1/3}T^{2/3} + \sqrt{\binom{n}{k-L}T})}$ **(This work)**
Degree d Polynomial	×	$R(S^*)$	$\min(\sqrt{n^dT}, \sqrt{n^kT})$ Chen et al. (2021)	$\min(\sqrt{n^dT}, \sqrt{n^kT})$ Chen et al. (2021)
Submodular+monotone (This work)	✓	$R_{\textbf{gr}}$	$\boldsymbol{\min_L(Ln^{1/3}T^{2/3} + \sqrt{\binom{n}{k-L}T})}$	$\boldsymbol{\min_i(L^{1/3}n^{1/3}T^{2/3} + \sqrt{\binom{n}{k-L}T})}$

Table 1: Best known regret bounds for combinatorial multiarmed bandits under different assumptions. By lemma 1.1 our upperbound can also be stated for $R_{1-e^{-1}}$. We note that our lower bound proven for the stochastic setting immediately applies to the adversarial setting in the table.

This notion of regret captures the fact that if the algorithm plays a set in $\mathcal{S}^{k,\boldsymbol{\epsilon}}$ then they may be incurring up to $\mathbf{1}^T\boldsymbol{\epsilon}$ extra regret. Note that when $\boldsymbol{\epsilon} = \mathbf{0}$ achieves the minimum (which can happen if the “gaps” between the greedily added element and all other elements at each stage is large) then this notion of regret is relative to the greedy set constructed in the noiseless setting.

The definition of regret in (3) is not novel to our paper. This notion is implicitly used in Streeter and Golovin (2007) in the proofs of Lemma 3 for the full-feedback setting and Theorem 13 for the bandit feedback setting, Nie et al. (2022) in Theorem 4.1, Pedramfar and Aggarwal (2023) in Theorem 1, Niazadeh et al. (2023) in Theorem 2 for the full-feedback setting and Theorem 4 for bandit feedback, and Nie et al. (2023) in Theorem 1. However, readers of these papers will note that they report their results not in terms of $R_{\textbf{gr}}$, but α-Regret: for an $\alpha \in [0, 1]$, define α-regret by, $R_\alpha := \sum_{t=1}^{T} \alpha f(S^*) - f(S_t)$ where $S^* := \arg\max_{|S|\leq k} f(S)$. Using Lemma 1, one immediately has that $R_\alpha \leq R_{\textbf{gr}}$ for $\alpha = (1 - e^{-1})$. Thus, an upper bound on (3) immediately results in an upper bound on R_α, which is precisely what previous works exploit to obtain their upper bounds on R_α.

To summarize: all the analyses of these previous works concentrate on showing an upper bound on $R_{\textbf{gr}}$, and only at the last step argue that $R_\alpha \leq R_{\textbf{gr}}$, and report an upper bound on R_α. But R_α can be a very loose lower bound on $R_{\textbf{gr}}$! For instance, when the function is modular (the inequalities of submodularity are tight), and the gap between the best set and worst set is equal to $\Delta < e^{-1}$, then a random selection algorithm would get zero or even negative R_α regret, while $R_{\textbf{gr}}$ would be linear ΔT, which is more natural. Thus, in studying regret against approximations attained by an offline step-wise greedy procedure, R_{gr} can be a more appropriate measure than R_α

1.2 Related Work

There has been several works on combinatorial multi-armed bandits with submodular assumptions and different feedback assumptions. Table 1 summarizes of the most relevant results as well as the results of this paper. For monotonic submodular maximization specifically, previous work use Lemma 1.1 with appropriate $\boldsymbol{\epsilon}$ to prove an upper bound on expected R_α-regret when the greedy result with perfect information gives an α-approximation of the actual maximum value.

Stochastic In the stochastic setting, when the expected reward function is submodular and monotonic, Nie et al. (2022) proposed an explore-then-commit algorithm with full-bandit feedback that achieves $R_{\textbf{gr}} = \mathcal{O}(k^{4/3}T^{2/3}n^{1/3})^2$. Recently, Pedramfar and Aggarwal (2023) showed with the same explore-then-commit algorithm with different parameters, $R_{\textbf{gr}} = \mathcal{O}(kn^{1/3}T^{2/3} + kn^{2/3}T^{1/3}d)$ is possible with delay feedback parameter of d. Without the monotonicity, Fourati et al. (2023) achieves $R_\alpha = \mathcal{O}(nT^{2/3})$ with bandit feedback for $\alpha = 1/2$. There have also been several works in the semi-bandit feedback setting (Wen et al. (2017), Zhu et al. (2021)), and others such as getting the marginal gain of each element after each query.

Adversarial In the adversarial setting, the environment chooses an arbitrary sequence of monotone submodular functions $\{f_1, \ldots, f_T\}$, and the goal is to minimize regret against an approximation of the reward of the best set in hindsight (Golovin et al. (2014), Harvey et al. (2020), Streeter et al. (2009),

[2] Most previous works, Nie et al. (2022); Pedramfar and Aggarwal (2023), state their result in terms of R_α however, a careful analysis of the proofs of their main regret bounds show a stronger result in terms of $R_{\textbf{gr}}$.

Wan et al. (2023)). Streeter and Golovin (2008) showed $\mathcal{O}(k\sqrt{Tn\log n})$ $R_{(1-e^{-1})}$-regret is possible with partially transparent feedback(where after each round, $f(S^{(i)})$ for all i is revealed instead of only $f(S^{(k)})$) and $\mathcal{O}(kn^{1/3}T^{2/3})$ $R_{(1-e^{-1})}$-regret for the bandit-feedback setting. Niazadeh et al. (2023) proposed a generalized algorithm with $\tilde{\mathcal{O}}(kn^{2/3}T^{2/3})$ $R_{(1-e^{-1})}$-regret with full bandit feedback, and showed all explore-then-commit greedy algorithms have $\Omega(T^{2/3})$ regret, when applied to our setting. Without the monotone assumption, Niazadeh et al. (2023) gets $\mathcal{O}(nT^{2/3})$ $R_{(1/2)}$-regret with bandit feedback. The upper-bound results in the adversarial setting doesn't naturally lead to results in the stochastic setting as the function is submodular and monotone only in expectation in the stochastic setting.

Continuous Submodular There are several works on online maximization of the continuous extensions of submodular set functions to a compact subspace such as Lovász and multilinear extensions(Bach (2019), Feldman and Karbasi (2020)). With a stronger assumption of DR-submodularity, it's possible to achieve higher approximation ratio guarantees and lower regret bounds (Bian et al. (2017a), Bian et al. (2017b), Sadeghi et al. (2021)). Wan et al. (2023) uses multilinear extension to achieve $O(T^{2/3})$ $R_{(1-e^{-1})}$-regret for adversarial submodular maximization with partition matroid constraint.

Low-degree polynomial In general reward functions without the submodular assumption, Chen et al. (2021) showed if the reward function is a d-degree polynomial, $\Theta\big(\min(\sqrt{n^dT}, \sqrt{n^kT})\big)$ regret is optimal.

2 LOWER BOUND

Theorem 2.1. *For any $n \geq 4$, $k \leq \lfloor n/3 \rfloor$, satisfying $512k^7n \leq T \in \mathbb{N}$, let $\mathcal{F}$ denote the set of submodular functions that are non-decreasing and bounded by $[0,1]$ for sets of size k or less, with $f(\emptyset) = 0$. Then*

$$\inf_{\mathsf{Alg}} \sup_{f\in\mathcal{F}} \mathbb{E}[R_{\boldsymbol{gr}}] \geq \frac{1}{16}(k-i^*)^{1/3}T^{2/3}n^{1/3}e^{-8} + \frac{1}{4}T^{1/2}\sqrt{\binom{n-k}{i^*}}e^{-2}$$

where the infimum is over all randomized algorithms and the supremum is over the functions in $\mathcal{F}$, and $i^* \in [k]$ is the largest value satisfying $\frac{16}{n^2k^6}\binom{n-k}{i^*}^3 \leq T$.

The lowerbound is intuitively a mix of the greedy explore-then-commit algorithm for the first $k - i^*$ arms, and then a standard MAB algorithm between all superarms of cardinality k that include those elements. For small T (i.e. $T = \mathcal{O}(n^4)$) the regret would be $\Omega(k^{1/3}n^{1/3}T^{2/3})$, and for large T(i.e. $T = \Omega(n^{3k-2})$) the regret would be $\Omega(\binom{n}{k}^{1/2}T^{1/2})$. This lowerbound also immediately gives a lower bound for the adversarial setting where $\tilde{f}_i = f + \mathcal{N}(0,1)$ is the function chosen by the environment at time i.

Proof Sketch We construct a hard instance so that at each cardinality a single set gives an elevated reward. Focusing on $k = 2$ for illustration, the instance would be the following:

$$\mathbf{H}_0 := \begin{cases} f(\{i\}) = 1/2 & \text{if } i \in \{1\} \\ f(\{i\}) = 1/2 - \Delta & \text{if } i \in [n] \setminus \{1\} \\ f(\{i,j\}) = 3/4 & \text{if } (i,j) = (1,2) \\ f(\{i,j\}) = 3/4 - \Delta & \text{if } (i,j) \in \binom{[n]}{2} \setminus \{(1,2)\} \end{cases}$$

where Δ is the gap of the best set that we will tune based on T. Pulling any arm of cardinality less than 2 would incur $\Omega(1)$ regret, however, since there are only n such sets (compared to $\binom{n}{2}$ sets of size 2), pulling these simple arms give more information on the optimal set.

For a set of alternative instances, we choose a set of size k and elevate its reward by 2Δ. We also elevate every prefix set of a permutation of this set by 2Δ so that the new set can be found by a greedy algorithm. Again, for $k = 2$, and any $\{\hat{i}, \hat{j}\} \in [n]\backslash\{1, 2\}$

$$\mathbf{H}_{\widehat{i},\widehat{j}} := \begin{cases} f(\{i\}) = 1/2 & \text{if } i \in \{1\} \\ f(\{i\}) = 1/2 + \Delta & \text{if } i \in \{\widehat{i}\} \\ f(\{i\}) = 1/2 - \Delta & \text{if } i \in [n] \setminus \{1, \widehat{i}\} \\ f(\{i,j\}) = 3/4 & \text{if } (i,j) = (1,2) \\ f(\{i,j\}) = 3/4 + \Delta & \text{if } (i,j) = (\widehat{i}, \widehat{j}) \\ f(\{i,j\}) = 3/4 - \Delta & \text{Otherwise} \end{cases}$$

Note that, if $\Delta < \frac{1}{16}$ for the $k = 2$ instance, All the functions are submodular, as $f(\{a, b\}) - f(\{b\}) \leq \frac{1}{4} + 2\Delta \leq 1/2 - \Delta \leq f(\{a\}) - f(\{\phi\})$ for any $a, b \in [n]$.

For $\mathbf{H}_0$, if $\epsilon_i < \Delta$ for all $i \in [2]$, then $f_{\mathcal{H}_0}(S_{gr}^{2,\boldsymbol{\epsilon}}) = \frac{3}{4}$ as the noisy greedy finds the best arm, and otherwise $\mathbf{1}^T\boldsymbol{\epsilon} \geq \Delta$, so $\min_{\boldsymbol{\epsilon}\geq\mathbf{0}} f_{\mathbf{H}_0}(S_{gr}^{2,\boldsymbol{\epsilon}}) + \mathbf{1}^T\boldsymbol{\epsilon} = \frac{3}{4}$. Similarly, $\min_{\boldsymbol{\epsilon}\geq\mathbf{0}} f_{\mathbf{H}_{\widehat{i},\widehat{j}}}(S_{gr}^{2,\boldsymbol{\epsilon}}) + \mathbf{1}^T\boldsymbol{\epsilon} = \frac{3}{4} + \Delta$. So for these instances $R_{\mathbf{gr}} = R(S^*)$.

We show that if the KL divergence between an alternate instance and $\mathbf{H}_0$ is small, then the algorithm cannot distinguish between the two environments and the maximum regret of the two would be $\Omega(\Delta T)$. Let $\mathbb{P}_{\widehat{i},\widehat{j}}, \mathbb{E}_{\widehat{i},\widehat{j}}$ be the probability and expectation under $\mathbf{H}_{\widehat{i},\widehat{j}}$, respectively when executing some fixed algorithm with observations being corrupted by standard Gaussian noise. Then $KL(\mathbb{P}_0|\mathbb{P}_{\widehat{i},\widehat{j}}) = \frac{\Delta^2}{2}\big(\mathbb{E}_0[T_{\widehat{i}}] + 4\mathbb{E}_0[T_{\widehat{i},\widehat{j}}]\big)$ for $k = 2$, where T_S is the number of pulls of set S, and

$$\mathbb{E}_0[R_{\mathbf{gr}}] + \mathbb{E}_{\widehat{i},\widehat{j}}[R_{\mathbf{gr}}] \geq \tfrac{1}{2}\sum_{i=1}^{n} \mathbb{E}_0[T_i] + \tfrac{\Delta T}{2}\Big(\mathbb{P}_0(T_{1,2} \leq \tfrac{T}{2}) + \mathbb{P}_{\widehat{i},\widehat{j}}(T_{1,2} > \tfrac{T}{2})\Big)$$

$$\geq \tfrac{1}{2}\sum_{i=1}^{n} \mathbb{E}_0[T_i] + \tfrac{\Delta T}{4}\exp(-KL(\mathbb{P}_0|\mathbb{P}_{\widehat{i},\widehat{j}})) = \tfrac{1}{2}\sum_{i=1}^{n} \mathbb{E}_0[T_i] + \tfrac{\Delta T}{4}\exp\Big(-2\Delta^2\big(\mathbb{E}_0[T_{\widehat{i}}] + \mathbb{E}_0[T_{\widehat{i},\widehat{j}}]\big)\Big).$$

Since $\widehat{i}, \widehat{j}$ were arbitrary, the following Lemma shows that there exist a pair that are pulled for small number of times in expectation (see Lemma A.2 for general k).

Lemma 2.2. *There exists a pair $\widehat{i}, \widehat{j}$ such that*

$$\mathbb{E}_0[T_{\widehat{i}}] + \mathbb{E}_0[T_{\widehat{i},\widehat{j}}] \leq \frac{2\sum_i \mathbb{E}_0[T_i]}{n-2} + \frac{T}{\binom{n-2}{2}}$$

Proof. For a pair (i, j), define $Q_{(i,j)} := \mathbb{E}_0[T_i] + \mathbb{E}_0[T_{i,j}]$. Then the sum of this term for all pairs not equal to $1, 2$ would be

$$Q := \sum_{(i,j)\neq(1,2)} Q_{(i,j)} \leq (n-3)\sum_{i\neq(1,2))} \mathbb{E}_0[T_i] + \sum_{i,j\neq 1,2} \mathbb{E}_0[T_{i,j}] \leq (n-3)\sum_i \mathbb{E}_0[T_i] + T$$

Then by Pigeonhole principal there exist a pair $\widehat{i}, \widehat{j}$ such that

$$Q_{(\widehat{i},\widehat{j})} \leq \frac{Q}{\binom{n-2}{2}} \leq \frac{2}{n-2}\sum_i \mathbb{E}_0[T_i] + \frac{T}{\binom{n-2}{2}}$$

□

Using the lemma, for some $(\widehat{i}, \widehat{j})$, we have

$$\mathbb{E}_0[R_{\mathbf{gr}}] + \mathbb{E}_{\widehat{i},\widehat{j}}[R_{\mathbf{gr}}] \geq \frac{1}{2}\sum_{i=1}^{n} \mathbb{E}_0[T_i] + \tfrac{\Delta T}{4}\exp\Big(-2\Delta^2\big(\frac{2}{n-2}\sum_i \mathbb{E}_0[T_i] + \frac{T}{\binom{n-2}{2}}\big)\Big)$$

We choose an appropriate Δ based on value of i^*.

- If $i^* = 1$, then for $\Delta = T^{-1/3}n^{1/3}$, we have $\frac{2\Delta^2 T}{\binom{n-2}{2}} \leq 1$. So either the KL divergence is less than 2, then the regret is lowerbounded by $\Delta T e^{-2} = T^{2/3}n^{1/3}e^{-2}$, or for KL divergence to be larger than 2 we would have $\sum_i \mathbb{E}_0[T_i] \geq \frac{1}{4}T^{2/3}n^{1/3}$, which from the above equation shows the regret is $\Omega(T^{2/3}n^{1/3})$. This can be extended for expected value of pulls of each cardinality lower than $i^* + 1$ for general k.
- If $i^* = 2$, then it can be shown that the term $\frac{1}{2}\sum_{i=1}^n \mathbb{E}_0[T_i] + \frac{\Delta T}{4}\exp\Big(-2\Delta^2\big(\frac{2}{n-2}\sum_i \mathbb{E}_0[T_i] + (T - \sum_{i=1}^n \mathbb{E}_0[T_i])/\binom{n-2}{2}\big)\Big)$ with $\Delta = \sqrt{\binom{n-2}{2}/T}$ minimizes when $\sum_{i=1}^n \mathbb{E}_0[T_i] = 0$ i.e. zero single arm sets being pulled in expectation, so the regret would be $T^{1/2}\binom{n-2}{2}^{1/2}\exp(-2)$.

This shows that the expected regret is $\tilde{\Omega}(\min_i(i^{1/3}n^{1/3}T^{2/3} + \sqrt{\binom{n}{k-i}T}))$. The instance of general k, and the detailed proof is in appendix A.1. □

We define an algorithm to be in non-adaptive greedy error-threshold class against $R_{\mathbf{gr}}$ regret, if it selects $\epsilon'_1, \ldots, \epsilon'_k$ at the start only dependent on parameters T, n, k before any arm pulls, and minimizes regret against $f(S_{\mathbf{gr}}^{k,\boldsymbol{\epsilon}'}) + \mathbf{1}^T\boldsymbol{\epsilon}'$. All the algorithms from previous work in the literature fall within this restricted class, and with this extra assumption we can prove a stronger lower bound.

Theorem 2.3. *For any $n \geq 4$, $k \leq \lfloor n/3 \rfloor$, satisfying $512k^9 n \leq T \in \mathbb{N}$, let $\mathcal{F}$ denote the set of submodular functions that are non-decreasing and bounded by $[0,1]$ for sets of size k or less, with $f(\emptyset) = 0$. Then*

$$\inf_{\mathsf{Alg} \in \mathsf{NAET}} \sup_{f \in \mathcal{F}} \mathbb{E}[R_{\boldsymbol{gr}}] \geq \frac{1}{288}(k - i^*)T^{2/3}n^{1/3}e^{-10} + \frac{1}{4}T^{1/2}\sqrt{\binom{n-k}{i^*}}e^{-2}$$

where the infimum is over all randomized algorithms with non-adaptive greedy error threshold selection, and the supremum is over the functions in $\mathcal{F}$, and $i^* \in [k]$ is the largest value satisfying $\frac{16}{n^2k^6}\binom{n-k}{i^*}^3 \leq T$.

3 UCB UPPER BOUND

Algorithm 1 SUB-UCB algorithm for set bandits with cardinality constraints

```
Input: $T$, $m$, greedy stop level $l$
Initialization: $S^{(0)} = \emptyset$, $T_A = 0$ for all $A \subset [n]$
For each $a \in [n]$, pull $\{a\}$ exactly $m$ times and update $T_{\{a\}} \leftarrow m$. Update $t \leftarrow mn$.
for $i = 1, 2, \ldots, l$ do
    $U_a \leftarrow \infty$ for all $a \notin S^{(i-1)}$
    while $T_{S^{(i-1)} \cup \arg\max U_a} < m$ do
        Pull arm $S_t = S^{(i-1)} \cup \arg\max_a U_a$, observe $r_t$, and update $T_{S_t} \leftarrow T_{S_t} + 1$
        for each $a \notin S^{(i-1)}$ do
            $S_a \leftarrow S^{(i-1)} \cup \{a\}$
            $\hat{\mu}_{S_a} \leftarrow \frac{1}{T_{S_a}} \sum_{t: I_t = S_a} r_t$
            Compute UCB: $U_a = \hat{\mu}_{S_a} + \sqrt{\frac{8 \log t}{T_{S_a}}}$
        end for
        $t \leftarrow t + 1$
    end while
    Update the base set: $S^{(i)} \leftarrow S^{(i-1)} \cup \{a_i\}$ where $a_i := \arg\max_a U_a$
end for
while $t < T$ do
    Run UCB on all size $k$ super-arms $A$ where $S^{(l)} \in A$.
end while
```

A natural approach to minimizing regret is to take an Explore-Then-Commit strategy motivated by the greedy algorithm. Such an algorithm would be the following - proceed in k rounds. Set $S^0 = \emptyset$.

In round i pull each set in the collection $\{S^{i-1}\cup\{a\} : a\in[n]\setminus S^{i-1}\}$, m times. Use these samples to update our estimate $\hat{f}$ of f on these sets, and set $S^{(i)} \leftarrow \arg\max_{a\in[n]\setminus S^{i-1}} \hat{f}(S^{i-1}\cup\{a\})$. This approach has been pursued by existing works Nie et al. (2022), and with an appropriate choice of m results in $O(kn^{1/3}T^{2/3})$ regret.

The disadvantage of this approach is that it can not achieve the correct trade-off between $\sqrt{n^k T}$ and $kn^{1/3}T^{2/3}$ exhibited by the lower bound. Motivated by the statement of the lower bound, our algorithm SUB-UCB attempts to interpolate between these different regret regimes. The critical quantity is i^*. For the first $k-i^*$ cardinalities, our algorithm plays a UCB style strategy which more or less follows the ETC strategy described in the previous paragraph. After that, it defaults to a UCB algorithm on all subsets containing S^{k-i^*}, a total of $\binom{n-k+i^*}{i^*}$ possible arms.

Theorem 3.1. *For any $l\le k$,* SUB-UCB *guarantees*

$$\mathbb{E}[R_{\textbf{gr}}] \le (1+4\sqrt{2})lT^{2/3}n^{1/3}(\log T)^{1/3} + 65\sqrt{T\binom{n-k}{k-l}\log T} + \frac{32}{15}\binom{n-k}{k-l}$$

when $m = T^{2/3}n^{-2/3}\log T^{1/3}$.

Proof Sketch We show that for $\epsilon := 2\sqrt{2\log(2knT^2)/m}$, the greedy part of SUB-UCB with high probability adds an ϵ-optimal arm in each step. Defining event G to be $|\hat{\mu}_S - f(S)| \le \sqrt{2T_S\log(2knT^2)}$ for all iterations, we prove that this event is true with a probability of at least $1-\frac{1}{T}$.

On Event G, We show that an ϵ-good arm is selected at each step of the greedy algorithm for $\epsilon = 2\sqrt{\frac{2\log(2knT^2)}{m}}$. Let a be a sub-optimal arm with expected reward value more than $2\sqrt{\frac{2\log(2knT^2)}{m}}$ from the best arm in the i-th step i.e. $\Delta_{S^{(i)},a} := \max_{a'} f(S^{(i)}\cup\{a'\}) - f(S^{(i)}\cup\{a\}) \ge 2\sqrt{\frac{2\log(2knT^2)}{m}}$. Then if arm a is added in i-th step, we have $U_a(t)\ge U_{a^*}(t)\ge f(S^{(i))\cup\{a^*\}}$, and therefore,

$$U_a(t) - f(S^{(i)}\cup\{a\}) \ge \Delta_{S^{(i)},a} > 2\sqrt{\frac{2\log(2knT^2)}{m}},$$

so $\hat{\mu}_{S^{(i)}\cup\{a\}} - f(S^{(i)}\cup\{a\}) > \sqrt{\frac{2\log(2knT^2)}{m}}$. This is a contradiction with event G, so on event G such an arm cannot be selected. Lastly, we expand the regret of two stages. As UCB in the second part of the algorithm has the regret of $65\sqrt{T\binom{n-k}{k-l}\log T} + \frac{32}{15}\binom{n-k}{k-l}$ against the best arm containing $S^{(l)}$(see Lattimore and Szepesvari (2017)), it is an upper bound for the regret against the greedy solution were the first l steps select an ϵ-good arm, and the last $k-l$ steps select the best arm, so on event G the regret can be written against a set in $\mathcal{S}^{k,\boldsymbol{\epsilon}}$ where

$$\mathbf{1}^T\boldsymbol{\epsilon} = l\epsilon + (k-l)0 = 2l\sqrt{\frac{2\log(2knT^2)}{m}}.$$

Therefore, the expected regret $\mathbb{E}[R_{\textbf{gr}}]$ on event G can be written as

$$2Tl\sqrt{\frac{2\log(2knT^2)}{m}} + mn(k-i^*) + 65\sqrt{T\binom{n-k}{k-l}\log T} + \frac{32}{15}\binom{n-k}{k-l},$$

for any choice of m and l. So for $m = T^{2/3}n^{-2/3}\log^{1/3}(2knT^2)$ the above term becomes $\tilde{O}(lT^{2/3}n^{1/3} + \sqrt{T\binom{n}{k-l}})$. The detailed proof is in Appendix B □

4 EXPERIMENTS

For the experiments we compare SUB-UCB (l) for different greedy stop levels l, SUB-UCB ($k-i^*$) which selects the best stop level based on the regret analysis, the ETCG (explore-then-commit greedy) algorithm from Nie et al. (2022), and UCB on all size k arms. Each arm pull has a 1-Gaussian noise, with 50 trials for each setting. The expected reward functions are the following.

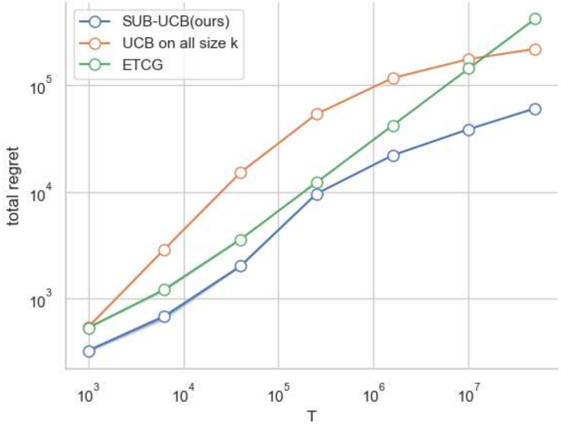

Figure 1: Regret comparison for weighted set cover with $n = 15$ and $k = 4$

Functions:

- The Unique greedy path hard instance i.e.

$$f(S) = \begin{cases} \sum_{i=1}^{|S|} \frac{1}{k+i} & S = \{1, \dots, |S|\} \\ \sum_{i=1}^{|S|} \frac{1}{k+i} + \frac{1}{100} & S = \{1, \dots, |S|\}. \end{cases}$$

  This function is inspired by the hard instance in the proof of our lower-bound. Note that this particular parameterization is submodular when $k \leq 7$, not for general k.

- Weighted set cover function i.e. $f_{\mathcal{C}}(S) = \sum_{C \in \mathcal{C}} w(C) \mathbf{1}\{S \cap C \neq \emptyset\}$ for a partition $\mathcal{C}$ of $[n]$ and weight function w on the partition. For $n = 15$ and $k = 4$, we use the partitions of size $5, 5, 4, 1$ with weights of $1/10, 1/10, 2/10, 6/10$ respectively.

Results: As illustrated in figure 1, we observe that our algorithm with the level selection of $k - i^*$ outperforms both ETCG and naive UCB on all size k arms, as it combines the advantages of greedy approach for small Ts and UCB on many super arms for large T. For smaller Ts compared to $\binom{n}{k}$, both SUB-UCB and ETCG outperform normal UCB as it doesn't have enough budget to find optimal sets of size k, so it gets linear regret(as the other two get $O(T^{2/3})$). However, as T becomes larger the reverse happens as $\binom{n}{k} T^{1/2}$ becomes smaller than $T^{2/3}$, but SUB-UCB adopts to T and continues to outperform the two until it converges with naive UCB for very large T.

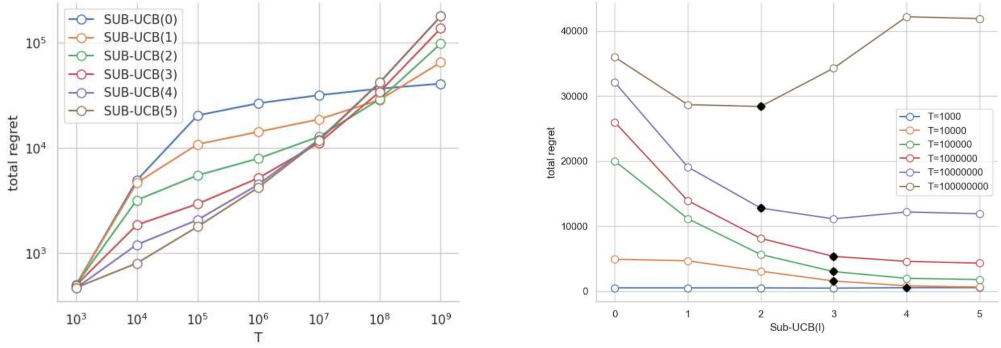

Figure 2: Comparison between all SUB-UCB greedy stop cardinality choices for the unique greedy path function with $n = 20$ and $k = 5$. The worst-case optimal stop cardinality $l = k - i^*$ is highlighted

In figure 2, we compare the performance of SUB-UCB for different choices of greedy stop cardinality, and observe that the best choice gradually decreases from k to 0 as T gets larger, and $k - i^*$ is a practical selection of the best stop cardinality before running the algorithm. Note that the defined stop level was chosen to minimize the worst-case bound on the regret, and if the gaps between arms on a particular instance are larger than the worst case, this stop level could be conservative. So $k - i^*$ is near the optimal stop level, and not the exact one as seen in these figures. Also, the empirical standard derivation is much smaller than $\mathcal{O}(T^{1/2})$ due to the regret symmetry of non-optimal sets at each cardinality, and it's not visible in the plots.

5 CONCLUSION

In this paper we showed that $\min_L(L^{1/3}T^{2/3}n^{1/3} + \sqrt{\binom{n}{k-L}T})$, ignoring logarithmic factors, is a lower bound on the regret against robust greedy solutions of stochastic submodular functions, and a stronger lower bound if the algorithm class is slightly restricted. We also matched this bound with an algorithm. This work is the first minimax lower bound for submodular bandits, and beyond closing the $k^{2/3}$ gap between the general lowerbound and upperbound, it remains open to prove similar minimax optimal bounds in settings with different types of constraint such as matroid, or in general, any offline-to-online greedy procedure that is robust to local noise (e.g. Non-monotonic submodular maximization where the greedy approach gets a $1/2$-approximation of the function, or DR-submodular optimization for the continuous setting which also has a $(1 - e^{-1})$-approximation).

Acknowledgements

This paper is based on work supported by Microsoft Grant for Customer Experience Innovation, NSF Award TRIPODS 2023239, and DMR-2308979.

References

Agarwal, M., Aggarwal, V., Quinn, C. J., and Umrawal, A. (2020). DART: aDaptive Accept RejecT for non-linear top-K subset identification. arXiv:2011.07687 [cs, stat].

Agarwal, M., Aggarwal, V., Quinn, C. J., and Umrawal, A. K. (2021). Stochastic Top-$K$ Subset Bandits with Linear Space and Non-Linear Feedback. arXiv:1811.11925 [cs, stat].

Bach, F. (2019). Submodular functions: from discrete to continuous domains. *Math. Program.*, 175(1–2):419–459.

Balcan, M.-F. and Harvey, N. J. A. (2011). Learning Submodular Functions.

Balkanski, E. and Singer, Y. (2018). The adaptive complexity of maximizing a submodular function. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 1138–1151, Los Angeles CA USA. ACM.

Bian, A. A., Buhmann, J. M., Krause, A., and Tschiatschek, S. (2019). Guarantees for Greedy Maximization of Non-submodular Functions with Applications. arXiv:1703.02100 [cs, math].

Bian, A. A., Levy, K. Y., Krause, A., and Buhmann, J. M. (2017a). Continuous dr-submodular maximization: structure and algorithms. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 486–496, Red Hook, NY, USA. Curran Associates Inc.

Bian, A. A., Mirzasoleiman, B., Buhmann, J., and Krause, A. (2017b). Guaranteed Non-convex Optimization: Submodular Maximization over Continuous Domains. In Singh, A. and Zhu, J., editors, *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pages 111–120. PMLR.

Chen, L., Hassani, H., and Karbasi, A. (2018). Online continuous submodular maximization. In Storkey, A. and Perez-Cruz, F., editors, *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, volume 84 of *Proceedings of Machine Learning Research*, pages 1896–1905. PMLR.

Chen, W., Hu, W., Li, F., Li, J., Liu, Y., and Lu, P. (2016a). Combinatorial Multi-Armed Bandit with General Reward Functions. In *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc.

Chen, W., Wang, Y., Yuan, Y., and Wang, Q. (2016b). Combinatorial Multi-Armed Bandit and Its Extension to Probabilistically Triggered Arms. arXiv:1407.8339 [cs].

Chen, X., Han, Y., and Wang, Y. (2021). Adversarial Combinatorial Bandits with General Non-linear Reward Functions. arXiv:2101.01301 [cs, stat].

Feige, U. (1998). A threshold of ln n for approximating set cover. *J. ACM*, 45(4):634–652.

Feldman, M. and Karbasi, A. (2020). Continuous submodular maximization: Beyond dr-submodularity. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H., editors, *Advances in Neural Information Processing Systems*, volume 33, pages 1404–1416. Curran Associates, Inc.

Fourati, F., Aggarwal, V., Quinn, C., and Alouini, M.-S. (2023). Randomized greedy learning for non-monotone stochastic submodular maximization under full-bandit feedback. In Ruiz, F., Dy, J., and van de Meent, J.-W., editors, *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, volume 206 of *Proceedings of Machine Learning Research*, pages 7455–7471. PMLR.

Goemans, M. X., Harvey, N. J. A., Iwata, S., and Mirrokni, V. (2009). Approximating Submodular Functions Everywhere. In *Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 535–544. Society for Industrial and Applied Mathematics.

Golovin, D., Krause, A., and Streeter, M. (2014). Online submodular maximization under a matroid constraint with application to learning assignments.

Hao, B., Lattimore, T., and Wang, M. (2021). High-Dimensional Sparse Linear Bandits. arXiv:2011.04020 [cs, math, stat].

Harvey, N., Liaw, C., and Soma, T. (2020). Improved Algorithms for Online Submodular Maximization via First-order Regret Bounds. In *Advances in Neural Information Processing Systems*, volume 33, pages 123–133. Curran Associates, Inc.

Hill, D. N., Nassif, H., Liu, Y., Iyer, A., and Vishwanathan, S. (2017). An efficient bandit algorithm for realtime multivariate optimization. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1813–1821.

Kaufmann, E., Cappé, O., and Garivier, A. (2016). On the Complexity of Best Arm Identification in Multi-Armed Bandit Models. arXiv:1407.4443 [cs, stat].

Kearns, M. and Singh, S. (2002). Near-optimal reinforcement learning in polynomial time. *Machine learning*, 49(2):209–232.

Krause, A. and Golovin, D. (2014). Submodular Function Maximization. In Bordeaux, L., Hamadi, Y., and Kohli, P., editors, *Tractability*, pages 71–104. Cambridge University Press, 1 edition.

Lattimore, T. and Szepesvari, C. (2017). Bandit algorithms.

Matsuoka, T., Ito, S., and Ohsaka, N. (2021). Tracking Regret Bounds for Online Submodular Optimization. In *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, pages 3421–3429. PMLR. ISSN: 2640-3498.

Nemhauser, G. L. and Wolsey, L. A. (1978). Submodular set functions, matroids and the greedy algorithm: Tight worst-case bounds and some generalizations of the rado-edmonds theorem. *Combinatorica*, 3(3-4):257–268.

Niazadeh, R., Golrezaei, N., Wang, J., Susan, F., and Badanidiyuru, A. (2023). Online Learning via Offline Greedy Algorithms: Applications in Market Design and Optimization. arXiv:2102.11050 [cs, math, stat].

Nie, G., Agarwal, M., Umrawal, A. K., Aggarwal, V., and Quinn, C. J. (2022). An Explore-then-Commit Algorithm for Submodular Maximization Under Full-bandit Feedback.

Nie, G., Nadew, Y. Y., Zhu, Y., Aggarwal, V., and Quinn, C. J. (2023). A framework for adapting offline algorithms to solve combinatorial multi-armed bandit problems with bandit feedback. In *Proceedings of the 40th International Conference on Machine Learning*, ICML'23. JMLR.org.

Pasteris, S. U., Rumi, A., Vitale, F., and Cesa-Bianchi, N. (2024). Sum-max submodular bandits. In Dasgupta, S., Mandt, S., and Li, Y., editors, *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pages 2323–2331. PMLR.

Pedramfar, M. and Aggarwal, V. (2023). Stochastic Submodular Bandits with Delayed Composite Anonymous Bandit Feedback. arXiv:2303.13604 [cs].

Roughgarden, T. and Wang, J. R. (2018). An optimal learning algorithm for online unconstrained submodular maximization. In Bubeck, S., Perchet, V., and Rigollet, P., editors, *Proceedings of the 31st Conference On Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pages 1307–1325. PMLR.

Sadeghi, O., Raut, P., and Fazel, M. (2021). Improved Regret Bounds for Online Submodular Maximization. arXiv:2106.07836 [cs, math, stat].

Simchowitz, M., Jamieson, K., and Recht, B. (2016). Best-of-K Bandits. arXiv:1603.02752 [cs, stat].

Singla, A., Tschiatschek, S., and Krause, A. (2016). Noisy submodular maximization via adaptive sampling with applications to crowdsourced image collection summarization. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, AAAI'16, page 2037–2043. AAAI Press.

Streeter, M. and Golovin, D. (2007). An Online Algorithm for Maximizing Submodular Functions:. Technical report, Defense Technical Information Center, Fort Belvoir, VA.

Streeter, M. and Golovin, D. (2008). An Online Algorithm for Maximizing Submodular Functions. In *Advances in Neural Information Processing Systems*, volume 21. Curran Associates, Inc.

Streeter, M., Golovin, D., and Krause, A. (2009). Online learning of assignments. In Bengio, Y., Schuurmans, D., Lafferty, J., Williams, C., and Culotta, A., editors, *Advances in Neural Information Processing Systems*, volume 22. Curran Associates, Inc.

Sviridenko, M., Vondrák, J., and Ward, J. (2014). Optimal approximation for submodular and supermodular optimization with bounded curvature. arXiv:1311.4728 [cs].

Svitkina, Z. and Fleischer, L. (2010). Submodular approximation: sampling-based algorithms and lower bounds. arXiv:0805.1071 [cs].

Wan, Z., Zhang, J., Chen, W., Sun, X., and Zhang, Z. (2023). Bandit multi-linear DR-submodular maximization and its applications on adversarial submodular bandits. In Krause, A., Brunskill, E., Cho, K., Engelhardt, B., Sabato, S., and Scarlett, J., editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 35491–35524. PMLR.

Wen, Z., Kveton, B., Valko, M., and Vaswani, S. (2017). Online influence maximization under independent cascade model with semi-bandit feedback. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 3026–3036, Red Hook, NY, USA. Curran Associates Inc.

Zhang, M., Chen, L., Hassani, H., and Karbasi, A. (2019). Online continuous submodular maximization: From full-information to bandit feedback. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.

Zhang, Q., Deng, Z., Chen, Z., Hu, H., and Yang, Y. (2022). Stochastic continuous submodular maximization: Boosting via non-oblivious function. In Chaudhuri, K., Jegelka, S., Song, L., Szepesvari, C., Niu, G., and Sabato, S., editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 26116–26134. PMLR.

Zhu, J., Wu, Q., Zhang, M., Zheng, R., and Li, K. (2021). Projection-free decentralized online learning for submodular maximization over time-varying networks. *Journal of Machine Learning Research*, 22(51):1–42.

A Lowerbound proofs

A.1 Proof of Theorem 2.1

For any $\{x_1, x_2, \dots, x_k\} \in \binom{[n]\setminus\{1,\dots,k\}}{k}$, define instance $\mathcal{H}_0, \mathcal{H}_{(x_1,\dots,x_k)}, \mathcal{H}_{(x_{i+1},\dots,x_k)}$ with reward functions as follows:

$$f_{\mathcal{H}_0}(S) := \begin{cases} H_{|S|+k} - H_k = \sum_{i=1}^{|S|} \frac{1}{k+i} & S = \{1, 2, \dots, |S|\} \\ H_{|S|+k} - H_k - \Delta & \text{Otherwise} \end{cases}$$

$$f_{\mathcal{H}_{(x_1,\dots,x_k)}}(S) := \begin{cases} H_{|S|+k} - H_k + \Delta & S = \{x_1, x_2, \dots, x_{|S|}\} \\ H_{|S|+k} - H_k & S = \{1, 2, \dots, |S|\} \\ H_{|S|+k} - H_k - \Delta & \text{Otherwise} \end{cases}$$

$$f_{\mathcal{H}_{(x_{i+1},\dots,x_k)}}(S) := \begin{cases} H_{|S|+k} - H_k + \Delta & S = \{1, \dots, i, x_{i+1}, \dots, x_{|S|}\} \\ H_{|S|+k} - H_k & S = \{1, 2, \dots, |S|\} \\ H_{|S|+k} - H_k - \Delta & \text{Otherwise} \end{cases}$$

where $H_n = \sum_{k=1}^{n} \frac{1}{k}$ is the n-th harmonic number.

Lemma A.1. *If $\Delta \leq (1/8k^2)$ then $\mathcal{H}_0$ and $\mathcal{H}_{(x_1,\dots,x_k)}$ are submodular.*

Proof. for any $S \subsetneq T \subset [n]$ where $|T| < k$ (the function is only defined on sets of cardinality at most k) and $x \notin T$ we have to show $f(S+x) - f(S) \geq f(T+x) - f(T)$.

$$f(T+x) - f(T) \leq \frac{1}{|T|+k+1} + 2\Delta \leq \frac{1}{|T|+k+1} + \frac{1}{4k^2} \leq \frac{1}{|T|+k} - \frac{1}{4k^2} \leq \frac{1}{|T|+k} - 2\Delta$$
$$\leq \frac{1}{|S|+1+k} - 2\Delta \leq f(S+x) - f(S)$$

□

For $\mathcal{H}_0$ if $\epsilon_i < \Delta$ at each step i of the greedy arm selection, then $S_{\mathbf{gr}}^{k,\boldsymbol{\epsilon}} = \{1, \dots, k\}$, otherwise $f_{\mathcal{H}_0}(S_{\mathbf{gr}}^{k,\boldsymbol{\epsilon}}) + \mathbf{1}^T\boldsymbol{\epsilon} \geq H_{2k} - H_k + \Delta - \Delta = H_{2k} - H_k = f_{\mathcal{H}_0}(\{1, \dots, k\})$. So $\min_{\boldsymbol{\epsilon}} f_{\mathcal{H}_0}(S_{\mathbf{gr}}^{k,\boldsymbol{\epsilon}}) + \mathbf{1}^T\boldsymbol{\epsilon} = f_{\mathcal{H}_0}(\{1, \dots, k\})$. This means that we can compute our regret against $f_{\mathcal{H}_0}(\{1, \dots, k\})$. Similarly, $\min_{\boldsymbol{\epsilon}} f_{\mathcal{H}_{(x_1,\dots,x_k)}}(S_{\mathbf{gr}}^{k,\epsilon}) + \mathbf{1}^T\boldsymbol{\epsilon} = H_{2k} - H_k + \Delta = f_{\mathcal{H}_{(x_1,\dots,x_k)}}(\{x_1, \dots, x_k\})$ showing that we can compute our regret against $\{x_1, \cdots, x_k\}$.

Let $\mathbb{E}_0$ and $\mathbb{E}_{(x_1,\dots,x_k)}$ denote the probability law under $\mathcal{H}_0$ and $\mathcal{H}_{(x_1,\dots,x_k)}$, respectively. For any $S \subset [n]$ let T_S denote the random variable describing the number of time the set S is played by a policy π. Define $T_i := \sum_{S\subset[n]:|S|=i} T_S$.

Then by the definition of $\mathcal{H}_0$ we have

$$\begin{aligned}
\mathbb{E}_0[R_{\mathbf{gr}}] &\geq \sum_{i=1}^{k-1} (f_{\mathcal{H}_0}(1, \dots, k) - \max_{S:|S|=i} f_{\mathcal{H}_0}(S))\mathbb{E}_0[T_i] + \sum_{S:|S|=k} (f_{\mathcal{H}_0}(\{1, \dots, k\}) - f_{\mathcal{H}_0}(S))\mathbb{E}_0[T_S] \\
&\geq \sum_{i=1}^{k-1} \Big(\sum_{j=i+1}^{k} 1/(k+j) \Big) \mathbb{E}_0[T_i] + \Delta \sum_{\{y_1,\dots,y_k\}\neq\{1,\dots,k\}} \mathbb{E}_0[T_{\{y_1,\dots,y_k\}}] \\
&\geq \sum_{i=1}^{k-1} \frac{k-i}{2k} \mathbb{E}_0[T_i] + \frac{\Delta T}{2} \mathbb{P}_0(T_{\{1,\dots,k\}} \leq T/2)
\end{aligned}$$

Similarly for $\mathcal{H}_{(x_1,\dots,x_k)}$ we have

$$\begin{aligned}
&\mathbb{E}_{\{x_1,\dots,x_k\}}[R_{\mathbf{gr}}] \\
&\geq \sum_{i=1}^{k-1}(f_{\mathcal{H}_{(x_1,\dots,x_k)}}(\{x_1,\dots,x_k\}) - \max_{|S|=i} f_{\mathcal{H}_{(x_1,\dots,x_k)}}(S))\mathbb{E}_{(x_1,\dots,x_k)}[T_i] \\
&+ \sum_{S:|S|=k} (f_{\mathcal{H}_{(x_1,\dots,x_k)}}(\{x_1,\dots,x_k\}) - f_{\mathcal{H}_{(x_1,\dots,x_k)}}(S))\mathbb{E}_0[T_S] \\
&\geq \sum_{i}^{k-1}(\sum_{j=i+1}^{k} 1/(k+j))\mathbb{E}_{\{x_1,\dots,x_k\}}[T_i] + \Delta \sum_{\{y_1,\dots,y_k\}\neq\{x_1,\dots,x_k\}} \mathbb{E}_{\{x_1,\dots,x_k\}}[T_{\{y_1,\dots,y_k\}}] \\
&\geq \frac{\Delta T}{2}\mathbb{P}_{\{x_1,\dots,x_k\}}(T_{\{1,\dots,k\}} > T/2).
\end{aligned}$$

Lemma A.2. *For any $i \leq k$ here exist a sequence $(x_i,\dots,x_k)$, where*

$$\begin{aligned}
&\sum_{j=i}^{k} \mathbb{E}_0[T_{\{1,\dots,i-1,x_i,\dots,x_j\}}] \\
&\leq \frac{1}{n-k}\mathbb{E}_0[T_i] + \frac{2}{(n-k)(n-k-1)}\mathbb{E}_0[T_{i+1}] + \frac{4}{(n-k)(n-k-1)}\sum_{j=i+2}^{k-1}\frac{k-j}{2k}\mathbb{E}_0[T_j] + \frac{T}{\binom{n-k}{k-i+1}}.
\end{aligned}$$

Proof. For $i \leq k$ and a sequence $(x_i, ..., x_k)$, define $Q_{(x_i,\dots,x_k)} := \sum_{j=i}^{k} \mathbb{E}_0[T_{\{1,\dots,i-1,x_i,\dots,x_j\}}]$. Then we have

$$Q := \sum_{(x_i,\dots,x_k)\neq(i,\dots,k)} Q_{(x_i,\dots,x_k)} \leq \sum_{j=i}^{k-1} \frac{(n-k-j+i-1)!(j-i+1)!}{(n-2k+i-1)!}\mathbb{E}_0[T_j] + ((k-i+1)!)\mathbb{E}_0[T_k].$$

Then by Pigeonhole principle, the exists a sequence $(x_i,\dots,x_k)$ such that

$$\begin{aligned}
Q_{(x_i,\dots,x_k)} &\leq \frac{Q}{\frac{(n-k)!}{(n-2k+i-1)!}} \\
&\leq \sum_{j=i}^{k-1}\frac{(n-k-j+i-1)!(j-i+1)!}{(n-k)!}\mathbb{E}_0[T_j] + \frac{(n-2k+i-1)!(k-i+1)!}{(n-k)!}\mathbb{E}[T_k] \\
&\leq \frac{1}{n-k}\mathbb{E}_0[T_i] + \frac{1}{(n-k)(n-k-1)}\sum_{j=i+1}^{k-1}\frac{(j-i)(j-i-1)}{\binom{n-k-2}{j-i-2}}\mathbb{E}[T_j] + \frac{1}{\binom{n-k}{k-i+1}}\mathbb{E}[T_k] \\
&\leq \frac{1}{n-k}\mathbb{E}_0[T_i] + \frac{2}{(n-k)(n-k-1)}\mathbb{E}_0[T_{i+1}] \\
&\quad + \frac{4}{(n-k)(n-k-1)}\sum_{j=i+2}^{k-1}\frac{k-j}{2k}\mathbb{E}_0[T_j] + \frac{T}{\binom{n-k}{k-i+1}}.
\end{aligned}$$

□

Lemma A.3. *For $\mathcal{H}_0$ and $\mathcal{H}_{(x_1,\dots,x_k)}$ defined above, we have*

$$KL(\mathbb{P}_0|\mathbb{P}_{\{x_i,\dots,x_k\}}) = 2\Delta^2\sum_{j=i}^{k-1}\mathbb{E}_0[T_{1,\dots,i-1,x_i,\dots,x_j}]$$

Proof.

$$
\begin{aligned}
KL(\mathbb{P}_0|\mathbb{P}_{\{x_i,\dots,x_k\}}) &= \sum_{S:|S|\le k} \mathbb{E}_0[T_S] KL(P_0(S)|P_{\{x_i,\dots,x_k\}}(S)) \\
&\qquad\qquad\qquad \text{(lemma 15.1 in Lattimore and Szepesvari (2017))} \\
&= \sum_{j=i}^{k} 2\Delta^2 \mathbb{E}_0[T_{1,\dots,i-1,x_i,\dots,x_j}]
\end{aligned}
$$

where $P_0(S) = \mathcal{N}(f_{\mathcal{H}_0}(S), 1)$ and $P_{\{x_i,\dots,x_k\}}(S) = \mathcal{N}(f_{\mathcal{H}_{(x_i,\dots,x_k)}}(S), 1)$ are the reward distributions of arm S in $\mathcal{H}_0$ and $\mathcal{H}_{(x_i,\dots,X_k)}$ respectively.

□

Using two above lemmas, we have,

$$
\begin{aligned}
&2\max\Big(\mathbb{E}_0[R_{\mathbf{gr}}], \max_{1\le i\le k,(x_i,\dots,x_k)\neq(i,\dots,k)} \mathbb{E}_{\{x_i,\dots,x_k\}}[R_{\mathbf{gr}}]\Big) \\
&\ge \max_{1\le i\le k,(x_i,\dots,x_k)\neq(i,\dots,k)} E_0[R_{\mathbf{gr}}] + \mathbb{E}_{\{x_i,\dots,x_k\}}[R_{\mathbf{gr}}] \\
&\ge \max_{1\le i\le k,(x_i,\dots,x_k)\neq(i,\dots,k)} \frac{\Delta T}{2}\Big(\mathbb{P}_0(T_{\{1,\dots,k\}} \le T/2) + \mathbb{P}_{\{x_i,\dots,x_k\}}(T_{\{1,\dots,k\}} > T/2)\Big) \\
&\ge \max_{1\le i\le k,(x_i,\dots,x_k)\neq(i,\dots,k)} \frac{\Delta T}{2} \exp(-KL(\mathbb{P}_0|\mathbb{P}_{\{x_i,\dots,x_k\}})) \\
&\qquad\qquad\qquad \text{(Using Pinsker's Inequality Lattimore and Szepesvari (2017))} \\
&\ge \max_{1\le i\le k,(x_i,\dots,x_k)\neq(i,\dots,k)} \frac{\Delta T}{2} \exp\Big(-2\Delta^2 \sum_{j=i}^{k} \mathbb{E}_0[T_{\{1,\dots,i-1,x_i,\dots,x_j\}}]\Big) \qquad \text{(Using lemma A.3)} \\
&\ge \frac{\Delta T}{2} \max_{1\le i\le k} \exp\Big(-2\Delta^2\big(\frac{1}{n-k}\mathbb{E}_0[T_i] + \frac{2}{(n-k)(n-k-1)}\mathbb{E}_0[T_{i+1}] \\
&\quad + \frac{4}{(n-k)(n-k-1)} \sum_{j=i+2}^{k-1} \frac{k-j}{2k}\mathbb{E}_0[T_j] + \frac{T}{\binom{n-k}{k-i+1}}\big)\Big) \qquad \text{(Using Lemma A.2)} \\
&\ge \max_{1\le i\le k} \frac{1}{2}(k-i^*)^{1/3} T^{2/3} n^{1/3} \exp\Big(-2T^{-2/3}(k-i^*)^{2/3} n^{2/3}\big(\frac{1}{n-k}\mathbb{E}_0[T_i] \\
&\quad + \frac{2}{(n-k)(n-k-1)}\mathbb{E}_0[T_{i+1}] + \frac{4}{(n-k)(n-k-1)} \sum_{j=i+2}^{k-1} \frac{k-j}{2k}\mathbb{E}_0[T_j] + \frac{T}{\binom{n-k}{k-i+1}}))\Big) \\
&\qquad\qquad\qquad \text{(Setting } \Delta := ((k-i^*)n/T)^{1/3}\text{)}
\end{aligned}
$$

For $1 \le i \le k - i^* + 1$, $\frac{n^{2/3}T^{1/3}}{\binom{n-k}{k-i+1}} \le 1$ by definition of i^*; so either the maximum regret is larger than $\frac{1}{4}T^{2/3}(k-i^*)^{1/3}n^{1/3}\exp(-8)$, which proves the theorem, or $\frac{1}{n-k}\mathbb{E}_0[T_i] + \frac{2}{(n-k)(n-k-1)}\mathbb{E}_0[T_{i+1}] + \frac{4}{(n-k)(n-k-1)}\sum_{j=i+2}^{k-1}\frac{k-j}{2k}\mathbb{E}_0[T_j] \ge 3(1/\Delta^2)$. If the third term is larger than $1/\Delta^2$, then $\sum_{j=i+2}^{k-1}\frac{k-j}{2k}\mathbb{E}_0[T_j] \ge \frac{1}{16}\frac{n}{(k-i^*)^{2/3}}T^{2/3}n^{1/3}$ which proves the lowerbound as $\frac{n}{(k-i^*)^{2/3}} \ge (k-i^*)^{1/3}$. Therefore, the only remaining case is that either the first or second term is $\ge 1/\Delta^2$. This means that for $1 \le i \le k - i^* + 1$, either $\mathbb{E}_0[T_i] \ge \frac{1}{4}(k-i^*)^{-2/3}T^{2/3}n^{1/3}$ or $\mathbb{E}_0[T_{i+1}] \ge \frac{1}{8}(n-k-1)(k-i^*)^{-2/3}T^{2/3}n^{1/3} \ge \frac{1}{4}(k-i^*)^{-2/3}T^{2/3}n^{1/3}$. Therefore, for at least half of the $1 \le i \le k - i^* + 1$, $\mathbb{E}_0[T_i] \ge \frac{1}{4}(k-i^*)^{-2/3}T^{2/3}n^{1/3}$, and

$$
\mathbb{E}_0[R_{\mathbf{gr}}] \ge \sum_{j=1}^{k-i^*+1} \frac{k-j}{2k}\mathbb{E}_0[T_j] \ge \frac{1}{8}(k-i^*)^{1/3}T^{2/3}n^{1/3}.
$$

Note that since $T \ge 512k^7n$, we have $\Delta \le (kn/T)^{1/3} \le \frac{1}{8k^2}$, so the functions with this selection of Δ are submodular.

We now lower bound the regret in a different way. Let $\lambda := \frac{\sum_{j=k-i^*+1}^{k-1} \frac{k-i}{2k}\mathbb{E}_0[T_i]}{T}$, then $\lambda \leq 1$, and using lemma A.2 we have that there exists a selection of $(x_i, \dots, x_k)$ such that,

$$\sum_{j=i}^{k} \mathbb{E}_0[T_{\{1,\dots,i-1,x_i,\dots,x_j\}}]$$

$$\leq \frac{1}{n-k}\mathbb{E}_0[T_i] + \frac{2}{(n-k)(n-k-1)}\mathbb{E}_0[T_{i+1}] + \frac{4}{(n-k)(n-k-1)}\sum_{j=i+2}^{k-1}\frac{k-j}{2k}\mathbb{E}_0[T_j] + \frac{T}{\binom{n-k}{k-i+1}}$$

$$\leq \frac{4}{n-k}\sum_{j=i}^{k-1}\frac{k-j}{2k}\mathbb{E}_0[T_j] + \frac{T}{\binom{n-k}{k-i+1}} = \frac{4}{(n-k)}\lambda T + \frac{T}{\binom{n-k}{k-i+1}}$$

So

$$\begin{aligned}
&2\max\Big(\mathbb{E}_0[R_{\mathbf{gr}}], \max_{1\leq i\leq k,(x_i,\dots,x_k)\neq(i,\dots,k)} \mathbb{E}_{\{x_i,\dots,x_k\}}[R_{\mathbf{gr}}]\Big)\\
&\geq \max_{1\leq i\leq k,(x_i,\dots,x_k)\neq(i,\dots,k)} E_0[R_{\mathbf{gr}}] + \mathbb{E}_{\{x_i,\dots,x_k\}}[R_{\mathbf{gr}}]\\
&\geq \min_{\lambda\in[0,1]} \max_{1\leq i\leq k,(x_i,\dots,x_k)\neq(i,\dots,k)} \lambda T + \frac{\Delta T}{2}\exp\Big(-2\Delta^2\sum_{j=i}^{k}\mathbb{E}_0[T_{\{1,\dots,i-1,x_i,\dots,x_j\}}]\Big)\\
&\geq \min_{\lambda\in[0,1]} \max_{1\leq i\leq k} \lambda T + \frac{\Delta T}{2}\exp\Big(-2\Delta^2(\frac{4}{n-k}\lambda T + \frac{T}{\binom{n-k}{k-i+1}})\Big)\\
&\geq \min_{\lambda\in[0,1]} \max_{1\leq i\leq k-i^*-1} \lambda T + \frac{1}{2}T^{1/2}\binom{n-k}{k-i+1}^{1/2}\exp\Big(-2\frac{4\lambda\binom{n-k}{k-i+1}}{(n-k)} - 2\Big)\\
&\hfill \text{(Setting } \Delta := (\tbinom{n-k}{k-i+1}/T)^{1/2}\text{)}\\
&\geq \frac{1}{2}T^{1/2}\binom{n-k}{i^*}^{1/2}e^{-2}
\end{aligned}$$

The last inequality holds as $\log(\frac{4T^{1/2}\binom{n-k}{k-i+1}^{3/2}}{(n-k)T}) \leq 0$, and the function relative to λ is convex, $\lambda = 0$ minimizes in the last inequality. Combining the two parts of the proof we have

$$\begin{aligned}
&\max\Big(\mathbb{E}_0[R_{\mathbf{gr}}], \max_{1\leq i\leq k,(x_i,\dots,x_k)\neq(i,\dots,k)} \mathbb{E}_{\{x_i,\dots,x_k\}}[R_{\mathbf{gr}}]\Big)\\
&\geq \max\big(\frac{1}{8}(k-i^*)^{1/3}T^{2/3}n^{1/3}e^{-8}, \frac{1}{2}T^{1/2}\binom{n-k}{i^*}^{1/2}e^{-2}\big)\\
&\geq \frac{1}{16}(k-i^*)^{1/3}T^{2/3}n^{1/3}e^{-8} + \frac{1}{4}T^{1/2}\binom{n-k}{i^*}^{1/2}e^{-2}
\end{aligned}$$

A.2 Proof of Theorem 2.3

We generalize the lowerbound distance of Theorem 2.1 by having the gap Δ_i in cardinality i. For any $\{x_1, x_2, \dots, x_k\} \in \binom{[n]\setminus\{1,\dots,k\}}{k}$, define instance $\mathcal{H}_0, \mathcal{H}_{(x_1,\dots,x_k)}, \mathcal{H}_{(x_{i+1},\dots,x_k)}$ with reward functions as follows:

$$f_{\mathcal{H}_0}(S) := \begin{cases} H_{|S|+k} - H_k = \sum_{i=1}^{|S|}\frac{1}{k+i} & S = \{1, 2, \dots, |S|\}\\ H_{|S|+k} - H_k - \Delta_{|S|} & \text{Otherwise}\end{cases}$$

$$f_{\mathcal{H}_{(x_1,\dots,x_k)}}(S) := \begin{cases} H_{|S|+k} - H_k + \Delta_{|S|} & S = \{x_1, x_2, \dots, x_{|S|}\}\\ H_{|S|+k} - H_k & S = \{1, 2, \dots, |S|\}\\ H_{|S|+k} - H_k - \Delta_{|S|} & \text{Otherwise}\end{cases}$$

$$f_{\mathcal{H}_{(x_{i+1},\dots,x_k)}}(S) := \begin{cases} H_{|S|+k} - H_k + \Delta_{|S|} & S = \{1, \dots, i, x_{i+1}, \dots, x_{|S|}\}\\ H_{|S|+k} - H_k & S = \{1, 2, \dots, |S|\}\\ H_{|S|+k} - H_k - \Delta_{|S|} & \text{Otherwise}\end{cases}$$

The KL divergance between reward distribution of two instances is similarly:

$$KL(\mathbb{P}_0|\mathbb{P}_{\{x_i,\dots,x_k\}}) = \sum_{j=i}^{k} 2\Delta_j^2 \mathbb{E}_0[T_{1,\dots,i-1,x_i,\dots,x_j}]$$

Lemma A.4. *For any $i \leq k$ here exist a sequence $(x_i, \dots, x_k)$, where*

$$\begin{aligned}\sum_{j=i}^{k} \Delta_j^2 \mathbb{E}_0[T_{\{1,\dots,i-1,x_i,\dots,x_j\}}] \leq{} & \frac{1}{n-k}\Delta_i^2 \mathbb{E}_0[T_i] + \frac{2}{(n-k)(n-k-1)}\Delta_{i+1}^2 \mathbb{E}_0[T_{i+1}] \\ & + \frac{6}{(n-k)(n-k-1)(n-k-2)}\Delta_{i+2}^2 \\ & + \frac{12}{(n-k)(n-k-1)(n-k-2)} \sum_{j=i+3}^{k-1} \frac{k-j}{2k}\Delta_j^2 \mathbb{E}_0[T_j] + \frac{\Delta_k^2 T}{\binom{n-k}{k-i+1}}.\end{aligned}$$

Proof. For $i \leq k$ and a sequence $(x_i, ..., x_k)$, define $Q_{(x_i,\dots,x_k)} := \sum_{j=i}^{k} \Delta_j^2 \mathbb{E}_0[T_{\{1,\dots,i-1,x_i,\dots,x_j\}}]$. Then we have

$$Q := \sum_{(x_i,\dots,x_k)\neq(i,\dots,k)} Q_{(x_i,\dots,x_k)} \leq \sum_{j=i}^{k-1} \frac{(n-k-j+i-1)!(j-i+1)!}{(n-2k+i-1)!}\Delta_j^2 \mathbb{E}_0[T_j] + ((k-i+1)!)\Delta_k^2 \mathbb{E}_0[T_k].$$

Then by Pigeonhole principle, the exists a sequence $(x_i, \dots, x_k)$ such that

$$\begin{aligned}Q_{(x_i,\dots,x_k)} &\leq \frac{Q}{\frac{(n-k)!}{(n-2k+i-1)!}} \\ &\leq \sum_{j=i}^{k-1} \frac{(n-k-j+i-1)!(j-i+1)!}{(n-k)!}\Delta_j^2 \mathbb{E}_0[T_j] + \frac{(n-2k+i-1)!(k-i+1)!}{(n-k)!}\Delta_k^2 \mathbb{E}[T_k] \\ &\leq \frac{1}{n-k}\Delta_i^2 \mathbb{E}_0[T_i] + \frac{1}{(n-k)(n-k-1)} \sum_{j=i+1}^{k-1} \frac{(j-i)(j-i-1)}{\binom{n-k-2}{j-i-2}}\Delta_j^2 \mathbb{E}[T_j] + \frac{1}{\binom{n-k}{k-i+1}}\mathbb{E}[T_k] \\ &\leq \frac{1}{n-k}\Delta_i^2 \mathbb{E}_0[T_i] + \frac{2}{(n-k)(n-k-1)}\Delta_{i_1}^2 \mathbb{E}_0[T_{i+1}] \\ &\quad + \frac{6}{(n-k)(n-k-1)(n-k-2)}\Delta_{i+2}^2 \\ &\quad + \frac{12}{(n-k)(n-k-1)(n-k-2)} \sum_{j=i+3}^{k-1} \frac{k-j}{2k}\Delta_j^2 \mathbb{E}_0[T_j] + \frac{\Delta_k^2 T}{\binom{n-k}{k-i+1}}.\end{aligned}$$

□

We now assign Δ_i for lower cardinalities based on the value of Δ_k. If $\mathbf{1}^T \boldsymbol{\epsilon}' \leq 2\Delta_k$, For $i \leq k-1$, we assign $\Delta_i = \epsilon_i'$, so a greedy procedure with $\boldsymbol{\epsilon}'$ will retrieve the best set, hence $f_{\mathcal{H}_0}(S_{\mathbf{gr}}^{k,\boldsymbol{\epsilon}'}) + \mathbf{1}^T \boldsymbol{\epsilon}' \geq f_{\mathcal{H}_0}(\{1,\dots,k\})$ and $f_{\mathcal{H}_{(x_i,\dots,x_k)}}(S_{\mathbf{gr}}^{k,\boldsymbol{\epsilon}'}) + \mathbf{1}^T \boldsymbol{\epsilon}' \geq f_{\mathcal{H}_{(x_i,\dots,x_k)}}(\{1,\dots,i-1,x_i\dots,x_k\})$. Otherwise, since the gap of any set of size k and the best set is at most $2\Delta_k$ for both $\mathcal{H}_0$ and $\mathcal{H}_{(x_i,\dots,x_k)}$, $f_{\mathcal{H}_0}(S_{\mathbf{gr}}^{k,\boldsymbol{\epsilon}'}) + \mathbf{1}^T \boldsymbol{\epsilon}' \geq H_{2k} - H_k - \Delta_k + 2\Delta_k = f_{\mathcal{H}_0}(\{1,\dots,k\})$ and $f_{\mathcal{H}_{(x_i,\dots,x_k)}}(S_{\mathbf{gr}}^{k,\boldsymbol{\epsilon}'}) + \mathbf{1}^T \boldsymbol{\epsilon}' \geq H_{2k} - H_k - \Delta_k + 2\Delta_k \geq f_{\mathcal{H}_{(x_i,\dots,x_k)}}(\{1,\dots,i-1,x_i\dots,x_k\})$; so for $i \leq k-1$, and we assign $\Delta_i = \frac{\Delta_k}{k}$. Therefore, in both cases $R_{\mathbf{gr}} \geq R(S^*)$, and we give a lower bound for $R(S^*)$.

For the first part of the lower bound, we'll assign $\Delta_k = (k-i^*)(\frac{n}{T})^{1/3}$. Now similarly to proof of Theorem 2.1, we have

$$\begin{aligned}
&2\max\Big(\mathbb{E}_0[R_{\mathbf{gr}}], \max_{1\le i\le k,(x_i,\ldots,x_k)\neq(i,\ldots,k)} \mathbb{E}_{\{x_i,\ldots,x_k\}}[R_{\mathbf{gr}}]\Big)\\
&\ge \max_{1\le i\le k,(x_i,\ldots,x_k)\neq(i,\ldots,k)} \frac{\Delta_k T}{2}\exp\Big(-2\sum_{j=i}^{k}\Delta_j^2\mathbb{E}_0[T_{\{1,\ldots,i-1,x_i,\ldots,x_j\}}]\Big)\\
&\ge \frac{\Delta_k T}{2}\max_{1\le i\le k}\exp\Big(-2(\frac{1}{n-k}\Delta_i^2\mathbb{E}_0[T_i]+\frac{2}{(n-k)(n-k-1)}\Delta_{i+1}^2\mathbb{E}_0[T_{i+1}]\\
&+\frac{6}{(n-k)(n-k-1)(n-k-2)}\Delta_{i+2}^2+\frac{12}{(n-k)(n-k-1)(n-k-2)}\sum_{j=i+3}^{k-1}\frac{k-j}{2k}\Delta_j^2\mathbb{E}_0[T_j]\\
&+\frac{\Delta_k^2 T}{\binom{n-k}{k-i+1}})\Big) \qquad \text{(Using Lemma A.4)}\\
&\ge \max_{1\le i\le k}\frac{1}{2}(k-i^*)T^{2/3}n^{1/3}\exp\Big(-2(\frac{1}{n-k}\Delta_i^2\mathbb{E}_0[T_i]+\frac{2}{(n-k)(n-k-1)}\Delta_{i+1}^2\mathbb{E}_0[T_{i+1}]\\
&+\frac{6}{(n-k)(n-k-1)(n-k-2)}\Delta_{i+2}^2+\frac{12}{(n-k)(n-k-1)(n-k-2)}\sum_{j=i+3}^{k-1}\frac{k-j}{2k}\Delta_j^2\mathbb{E}_0[T_j]\\
&+\frac{(k-i^*)^2n^{2/3}T^{1/3}}{\binom{n-k}{k-i+1}}))\Big) \qquad \text{(Setting } \Delta_k := ((k-i^*)^3n/T)^{1/3}\text{)}
\end{aligned}$$

For $1\le i\le k-i^*+1$, $\frac{(k-i^*)^2n^{2/3}T^{1/3}}{\binom{n-k}{k-i+1}} \le \frac{k^2n^{2/3}T^{1/3}}{\binom{n-k}{k-i+1}} \le 1$ by definition of i^*; so either the maximum regret is larger than $\frac{1}{4}T^{2/3}(k-i^*)n^{1/3}\exp(-10)$, which proves the theorem, or

$$\begin{aligned}
&\frac{1}{n-k}\Delta_i^2\mathbb{E}_0[T_i]+\frac{2}{(n-k)(n-k-1)}\Delta_{i+1}^2\mathbb{E}_0[T_{i+1}]+\frac{6}{(n-k)(n-k-1)(n-k-2)}\Delta_{i+2}^2\\
&+\frac{12}{(n-k)(n-k-1)(n-k-2)}\sum_{j=i+3}^{k-1}\frac{k-j}{2k}\Delta_j^2\mathbb{E}_0[T_j]\Delta_j^2\mathbb{E}_0[T_j]\ge 4
\end{aligned}$$

If the forth term is larger than 1, then

$$\begin{aligned}
\Delta_k^2\sum_{j=i+3}^{k-1}\frac{k-j}{2k}\mathbb{E}_0[T_j] &\ge \sum_{j=i+3}^{k-1}\frac{k-j}{2k}\Delta_j^2\mathbb{E}_0[T_j]\\
&\ge \frac{(n-k)(n-k-1)(n-k-2)}{12}\ge\frac{n^3}{96}
\end{aligned}$$

So $\sum_{j=i+3}^{k-1}\frac{k-j}{2k}\mathbb{E}_0[T_j]\ge\frac{n^3}{96}\frac{1}{\Delta_k^2}\ge\frac{1}{96}(k-i^*)n^{1/3}T^{2/3}$ which proves the lower bound.

Therefore, the only remaining case is that at least one of the first three terms is ≥ 1. This means that for $1\le i\le k-i^*+1$, either $\mathbb{E}_0[T_i]\ge\frac{n}{2\Delta_i^2}$, or $\mathbb{E}_0[T_{i+1}]\ge\frac{n}{4\Delta_{i+1}^2}(n-k-1)\ge\frac{n}{4\Delta_{i+1}^2}$, or $\mathbb{E}_0[T_{i+2}]\ge\frac{n}{12\Delta_{i+2}^2}(n-k-1)(n-k-2)\ge\frac{n}{12\Delta_{i+2}^2}$.

Therefore, for at least $1/3$ of the $1\le i\le k-i^*+1$, $\mathbb{E}_0[T_i]\ge\frac{n}{12\Delta_i^2}$. Let I be all cardinalities in which this inequality holds(so $|I|\ge\frac{k-i^*}{3}$); since $\sum_{i=1}^{k-1}\Delta_i\le 2\Delta_k$, using Lemma C.1, we have

$$\mathbb{E}_0[R_{\mathbf{gr}}]\ge\sum_{j=1}^{k-i^*+1}\frac{k-j}{2k}\mathbb{E}_0[T_j]\ge\sum_{j\in I}\frac{k-j}{2k}\frac{n}{12\Delta_j^2}\ge\frac{1}{288}(k-i^*)T^{2/3}n^{1/3}.$$

For the second part of the lower bound, using $\Delta_i\le 2\Delta_k$, we have

$$KL(\mathbb{P}_0|\mathbb{P}_{\{x_i,\ldots,x_k\}})\le 8\Delta_k^2\sum_{j=i}^{k-1}\mathbb{E}_0[T_{1,\ldots,i-1,x_i,\ldots,x_j}]$$

, and the rest of the argument follows the proof of 2.1.

B Proof of Theorem 3.1

Proof. We use the notation $l = k - i^*$ to match our lowerbound, however as $k - i^*$ is arbitrary, it can be used for any other choice of l as well. Define the event $G := \bigcap_{i=1}^{k} \bigcap_{a \in [n] \setminus S^{(i-1)}} \bigcap_{t=1}^{T} g_{i,a,t}$ where

$$g_{i,a,t} := \left\{ \Big| \sum_{s \le t : I_s = S^{(i-1)} \cup \{a\}} (r_s - f(S^{(i-1)} \cup \{a\})) \Big| \le \sqrt{2T_{S^{(i-1)} \cup \{a\}}(t) \log(2knT^2)} \right\}.$$

Now note that if X_s are i.i.d. sub-Gaussian random variables then

$$\begin{aligned}
\mathbb{P}(G^c) &\le \sum_{i=1}^{k} \mathbb{P}\Big(\bigcup_{a \in [n] \setminus S^{(i-1)}} \bigcup_{t=1}^{T} g_{i,a,t}^c \Big) \\
&= \sum_{i=1}^{k} \sum_{S \in \binom{[n]}{i-1}} \mathbb{P}\Big(\bigcup_{a \in [n] \setminus S} \bigcup_{t=1}^{T} g_{i,a,t}^c | S^{(i-1)} = S \Big) \mathbb{P}(S^{(i-1)} = S) \\
&\le \sum_{i=1}^{k} \sum_{S \in \binom{[n]}{i-1}} \sum_{a \in [n] \setminus S} \mathbb{P}\Big(\bigcup_{t=1}^{T} g_{i,a,t}^c | S^{(i-1)} = S \Big) \mathbb{P}(S^{(i-1)} = S) \\
&\le \sum_{i=1}^{k} \sum_{S \in \binom{[n]}{i-1}} \sum_{a \in [n] \setminus S} \mathbb{P}\Big(\bigcup_{t=1}^{T} \{ | \sum_{s=1}^{t} X_s | \ge \sqrt{2t \log(2knT^2)} \} \Big) \mathbb{P}(S^{(i-1)} = S) \\
&\le \sum_{i=1}^{k} \sum_{S \in \binom{[n]}{i-1}} \sum_{a \in [n] \setminus S} \sum_{t=1}^{T} \frac{1}{knT^2} \mathbb{P}(S^{(i-1)} = S) \le 1/T.
\end{aligned}$$

Let $\mathcal{E}_i$ be the event that the arm selected at the i-th step of the algorithm is within $2\sqrt{\frac{2\log(2knT^2)}{m}}$ of the best possible arm at that step, i.e.

$$\mathcal{E}_i = \left\{ \max_{a \notin S^{(i-1)}} f(S^{(i-1)} \cup \{a\}) - f(S^{(i)}) \le 2\sqrt{\frac{2\log(2knT^2)}{m}} \right\}.$$

We prove that on event G, $\cup_{i \in [k-i^*]} \mathcal{E}_i$ is true.

Let a be a sub optimal arm with value more than $2\sqrt{\frac{2\log(2knT^2)}{m}}$ from the best arm in the i-th step. That is, if $a^* := \arg\max_{a'} f(S^{(i)} \cup \{a'\})$ and $\Delta_{S^{(i)},a} := f(S^{(i)} \cup \{a^*\}) - f(S^{(i)} \cup \{a\})$, then assume that $\Delta_{S^{(i)},a} \ge 2\sqrt{\frac{2\log(2knT^2)}{m}}$. Then on event G and arm a being added in i-th step,

$$U_a(t) \ge U_{a^*}(t) \ge f(S^{(i)} \cup \{a^*\}) = f(S^{(i)} \cup \{a\}) + \Delta_{S^{(i)},a}$$

which implies

$$U_a(t) - f(S^{(i)} \cup \{a\}) \ge \Delta_{S^{(i)},a} > 2\sqrt{\frac{2\log(2knT^2)}{m}}.$$

But this implies that

$$\hat{\mu}_{S^{(i)} \cup \{a\}} - f(S^{(i)} \cup \{a\}) > \sqrt{\frac{2\log(2knT^2)}{m}}$$

which is a contradiction of event G. Thus, on event G such an arm cannot be selected.

As UCB in the second part of the algorithm has the regret of $65\sqrt{T\binom{n-k}{k-l}\log T}+\frac{32}{15}\binom{n-k}{k-l}$ against $S^{(k)}$ which is the best size k arm containing $S^{(k-i^*)}$ (see Lattimore and Szepesvari (2017)), on event G, it is an upper bound for the regret against the greedy solution were the first $k-i^*$ steps select an ϵ-good arm, and the last i^* steps select the best arm, so on event G the regret can be written against a set in $\mathcal{S}^{k,\boldsymbol{\epsilon}}$ where

$$\mathbf{1}^T\boldsymbol{\epsilon}=(k-i^*)\epsilon=2(k-i^*)\sqrt{\frac{2\log\left(2knT^2\right)}{m}}.$$

Therefore, we upper bound the regret relative to $f(S^k)+2(k-i^*)\sqrt{\frac{2\log(2knT^2)}{m}}$, as by Lemma 1.1 it's greater than $\frac{1}{c}(1-e^{-c})f(S^*)$. Let T_i be the set of times where we pulled a set of cardinality i. From the while loop condition in the algorithm, we have $|T_i|\leq\sum_{a\notin S^{(i-1)}}\min\left\{\frac{1}{\Delta^2_{S^{(i-1)},a}},m\right\}\leq(n+1-i)m$ for $i\leq k-i^*$. For $\epsilon=2\sqrt{\frac{2\log(2knT^2)}{m}}$, we have

$$\begin{aligned}
\mathbb{E}[R_{\mathbf{gr}}]&\leq\mathbb{P}[G^c]T+\mathbb{E}[R_{\mathbf{gr}}\mathbf{1}\{G\}]\leq\frac{1}{T}T+\mathbb{E}[R_{\mathbf{gr}}\mathbf{1}\{G\}]\\
&\leq 1+\sum_{i=1}^{k-i^*}\sum_{t\in T_i}(f(S^{(k)})+(k-i^*)\epsilon)-f(S^{(i-1)}\cup\{a_t\}))+\sum_{t\in T_k}(f(S^{(k)})+(k-i^*)\epsilon)-f(S_t)\\
&\leq 1+(k-i^*)\epsilon T+mn(k-i^*)f(S^{(k)})+\sum_{t\in T_k}f(S^{(k)})-f(S_t)\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad(f(S^{(i-1)}\cup\{a_t\}))\geq 0)\\
&\leq 2T(k-i^*)\sqrt{\frac{2\log(2knT^2)}{m}}+mn(k-i^*)+65\sqrt{T\binom{n}{i^*}}+\frac{32}{15}\frac{n-k}{i^*}+1\\
&\leq T^{2/3}n^{1/3}(k-i^*)(\log(2knT^2))^{1/3}+\sqrt{8}T^{2/3}n^{1/3}(k-i^*)(\log(2knT^2))^{1/3}\\
&\quad+65\sqrt{T\binom{n}{i^*}}+\frac{32}{15}\frac{n-k}{i^*}+1.\qquad\qquad(\text{Setting } m=T^{2/3}n^{-2/3}\log^{1/3}(2knT^2))
\end{aligned}$$

□

C Auxiliary Lemmas

Lemma C.1. *For any sequence of numbers $a_1,\dots,a_n$ bounded between $(0,1]$, If $\sum_i a_i\leq C\leq 1$, then*

$$\sum_{i=1}^{n}\frac{1}{a_i^2}\geq\frac{n^3}{C^2}$$

Proof. If there exists $j,k\in[n]$ such that $a_j<a_k$, then for a new sequence $a'_i=\begin{cases}a_i & i\notin\{j,k\}\\ \frac{a_j+a_k}{2} & i\in\{j,k\}\end{cases}$ we have

$$\begin{aligned}
\sum a_i^{-2}-\sum a_i'^{-2}&=a_j^{-2}+a_k^{-2}-2\frac{4}{(a_j+a_k)^2}\\
&=\frac{2+\overbrace{a_j^2a_k^{-2}+a_j^{-2}a_k^2}^{>2}+2(\overbrace{a_j^{-1}a_k+a_ja_k^{-1}}^{>2})-8}{a_j^2+a_k^2+2a_ja_k}>0
\end{aligned}$$

Therefore, the infimum value of $\sum a_i^{-2}$ over all such sequences is when all elements are equal, and

$$\sum_{i=1}^{n}\frac{1}{a_i^2}\geq n\big(\frac{n}{\sum a_i}\big)^2\geq\frac{n^3}{C^2}.$$

□