An Explore-then-Commit Algorithm for Submodular Maximization Under Full-bandit Feedback Supplementary Material

Guanyu Nie[1] **Mridul Agarwal**[2] **Abhishek Kumar Umrawal**[2] **Vaneet Aggarwal**[2] **Christopher John Quinn**[1]

[1]Computer Science Department, Iowa State University, Ames, Iowa, USA
[2]Purdue University, West Lafayette, Indiana, USA

1 PROOFS

We will separate the proof of Theorem 4.1 into two cases. The first case is for when the clean event $\mathcal{E}$ defined in Section 4 happens, which we will show in Lemma 1.2 happens with high probability. Under the clean event, we will prove important preliminary results, namely Lemma 1.3 and Corollary 1.4. These will establish that even though ETCG, using random rewards, may pick a different sequence of subsets than an offline greedy algorithm [Nemhauser et al., 1978] using a value oracle for the expected reward function f, ETCG's chosen set of size k will nonetheless be near-optimal. The second case is when the complementary event happens, which occurs with low probability.

This proof structure is analogous to the standard MAB proof for explore-then-commit strategies (see for instance, Section 1.2 in [Slivkins, 2019]). However, unlike for standard MAB problems, ETCG makes sequences of decisions during exploration. Furthermore, the combinatorial action space and non-linear reward function make the problem challenging. Even in the special setting of deterministic rewards, the standard MAB problem becomes trivial (finding the largest of n base arms) while maximizing a submodular function with a cardinality constraint is NP-hard [Nemhauser et al., 1978].

1.1 PRELIMINARY

We first introduce some new notations and lemmas that are useful in the analysis. Recall from Section 2 that for an action $S \in \mathcal{S}$, $f_t(S)$ denotes a (random) reward at time t, $f(S)$ denotes the expected value for action S, and $\bar{f}_t(S)$ denotes the empirical mean of rewards received from playing action S up to and including time t. In the following, we will drop the subscript t from the empirical mean, writing $\bar{f}(S)$ when it is clear from context that action S has been played m times. Also recall that $S^{(i)}$ denotes the set of size $i \in \{1, \dots, k\}$ chosen after finishing phase i, and by the greedy structure of Algorithm 1, $\emptyset = S^{(0)} \subset S^{(1)} \subset \cdots \subset S^{(k)}$. This sequence of subsets that ETCG picks *does not necessarily match* the sequence chosen by the offline greedy approximation [Nemhauser et al., 1978] using a value oracle for the expected reward function f. Even though ETCG may select a different sequence, we will later show in Lemma 1.3 that with high probability, ensures the expected marginal gain is not too small.

Now we define events that are important in our analysis. Recall that $\bar{f}(S^{(i-1)} \cup \{a\})$ is the empirical mean of the m rewards from playing action $S^{(i-1)} \cup \{a\}$ in phase i. For each subset $S^{(i-1)} \cup \{a\}$, the m rewards are i.i.d. with mean $f(S^{(i-1)} \cup \{a\})$ and bounded in $[0, 1]$. Thus, we can bound the deviation of the (unbiased) empirical mean $\bar{f}(S^{(i-1)} \cup \{a\})$ from the expected value $f(S^{(i-1)} \cup \{a\})$ for each action in $\mathcal{S}_i$. Specifically, we can use a two-sided Hoeffding bound for bounded variables.

Lemma 1.1 (Hoeffding's inequality)**.** *Let $X_1, \cdots, X_n$ be independent random variables bounded in the interval $[0, 1]$, and let $\bar{X}$ denote their empirical mean. Then we have for any $\epsilon > 0$,*

$$\mathbb{P}\left(\left|\bar{X} - \mathbb{E}[\bar{X}]\right| \geq \epsilon\right) \leq 2\exp\left(-2n\epsilon^2\right). \tag{1}$$

We will use Hoeffding's inequality to bound the probabilities of the empirical means $\bar{f}(S^{(i-1)} \cup \{a\})$ for all actions $S^{(i-1)} \cup \{a\} \in \mathcal{S}_i$ played in phase i. By Algorithm 1, each action will be played the same number of times, denoted by

Accepted for the 38th Conference on Uncertainty in Artificial Intelligence (UAI 2022).

m, so we consider bounding the probabilities of equal-sized confidence radii $\mathrm{rad} := \sqrt{2\log(T)/m}$ for all the actions $S^{(i-1)} \cup \{a\} \in \mathcal{S}_i$ played in phase i.

We consider the event that the empirical means of all actions played in phase i are concentrated around their statistical means within a radius rad. Denote this event as $\mathcal{E}_i$,

$$\mathcal{E}_i := \bigcap_{S\cup\{a\}\in\mathcal{S}_i} \left\{ \left|\bar{f}(S\cup\{a\}) - f(S\cup\{a\})\right| < \mathrm{rad} \right\}. \tag{2}$$

Define the *clean event* $\mathcal{E}$ to be the event that the empirical means of all actions played up to and including phase k are within rad of their corresponding statistical means:

$$\mathcal{E} := \mathcal{E}_1 \cap \cdots \cap \mathcal{E}_k. \tag{3}$$

Lemma 1.2. *The probability of the clean event $\mathcal{E}$ defined in* (3) *satisfies:*

$$\mathbb{P}(\mathcal{E}) \geq 1 - \frac{2nk}{T^4}.$$

Proof. We begin by breaking up the probability of the clean event $\mathcal{E}$ into conditional probabilities for the events $\{\mathcal{E}_i\}_{i=1}^k$ for each phase,

$$\begin{aligned}\mathbb{P}(\mathcal{E}) &= \mathbb{P}(\mathcal{E}_1 \cap \cdots \cap \mathcal{E}_k) \\ &= \prod_{i=1}^{k} \mathbb{P}(\mathcal{E}_i|\mathcal{E}_1, \ldots, \mathcal{E}_{i-1}). \end{aligned} \tag{4}$$

Recall that $\mathcal{E}_i$, defined in (2), is the event where the empirical means of all actions played in phase i were concentrated around their statistical means. Which actions are available in phase i, namely $\{S^{(i-1)} \cup \{a\}\}_{a\in\Omega\setminus S^{(i-1)}}$, depends on the action $S^{(i-1)}$ from the previous phase that had the highest empirical mean, which in turn is related to $\mathcal{E}_{i-1}$. Although we cannot directly evaluate (4), by conditioning on $S^{(i-1)}$ we will be able to obtain a bound on (4).

$$\begin{aligned}\mathbb{P}(\mathcal{E}_i|\mathcal{E}_1, \ldots, \mathcal{E}_{i-1}) &= \sum_{S\in\left\{S' \,\middle|\, S'\subseteq\Omega,\ |S'|=i-1\right\}} \mathbb{P}(S^{(i-1)} = S, \mathcal{E}_i|\mathcal{E}_1, \ldots, \mathcal{E}_{i-1}) && \text{(law of total probability)} \\ &= \sum_{S\in\left\{S' \,\middle|\, S'\subseteq\Omega,\ |S'|=i-1\right\}} \mathbb{P}(S^{(i-1)} = S|\mathcal{E}_1, \ldots, \mathcal{E}_{i-1}) \times \mathbb{P}(\mathcal{E}_i|S^{(i-1)} = S, \mathcal{E}_1, \ldots, \mathcal{E}_{i-1}) \\ &= \sum_{S\in\left\{S' \,\middle|\, S'\subseteq\Omega,\ |S'|=i-1\right\}} \mathbb{P}(S^{(i-1)} = S|\mathcal{E}_1, \ldots, \mathcal{E}_{i-1}) \times \mathbb{P}(\mathcal{E}_i|S^{(i-1)} = S), \end{aligned} \tag{5}$$

where (5) follows from rewards in phase i being conditionally independent of rewards from other phases, given the corresponding actions played during phase i.

We now focus on bounding $\mathbb{P}(\mathcal{E}_i|S^{(i-1)} = S)$. By conditioning on the set chosen in the previous phase, $S^{(i-1)} = S$, we know all the actions that will be played in the current phase i, $\{S^{(i-1)} \cup \{a\}\}_{a\in\Omega\setminus S^{(i-1)}}$. The rewards of all the actions are bounded in $[0, 1]$ and are conditionally independent (given the corresponding action).

Apply Lemma 1.1 to the empirical mean $\bar{f}(S^{(i-1)} \cup \{a\})$ of m rewards for action $S^{(i-1)} \cup \{a\}$ and choosing $\epsilon = \mathrm{rad} = \sqrt{2\log(T)/m}$ gives

$$\begin{aligned}\mathbb{P}\left[\left|\bar{f}(S^{(i-1)} \cup \{a\}) - f(S^{(i-1)} \cup \{a\})\right| \geq \mathrm{rad}\right] &\leq 2\exp\left(-2m\mathrm{rad}^2\right) \\ &= 2\exp\left(-2m(2\log(T)/m)\right) \\ &= 2\exp\left(-4\log(T)\right) \\ &= \frac{2}{T^4}. \end{aligned}$$

Thus, for any individual action $S^{(i-1)} \cup \{a\} \in \mathcal{S}_i$, we can bound the probability that its sample mean $\bar{f}(S^{(i-1)} \cup \{a\})$ is within a specified confidence radius (complementary of the event above) as

$$\begin{aligned}\mathbb{P}\left[\left|\bar{f}(S^{(i-1)} \cup \{a\}) - f(S^{(i-1)} \cup \{a\})\right| < \mathrm{rad}\right] &= 1 - \mathbb{P}\left[\left|\bar{f}(S^{(i-1)} \cup \{a\}) - f(S^{(i-1)} \cup \{a\})\right| \geq \mathrm{rad}\right] \\ &\geq 1 - \frac{2}{T^4}. \end{aligned} \tag{6}$$

We can then use (6) to bound $\mathbb{P}(\mathcal{E}_i | S^{(i-1)} = S)$ for any set $S \subset \Omega$ of $i-1$ arms.

$$\begin{aligned}\mathbb{P}(\mathcal{E}_i | S^{(i-1)} = S) &= \mathbb{P}\left[\bigcap_{a \in \Omega \setminus S^{(i-1)}} \left\{\left|\bar{f}(S^{(i-1)} \cup \{a\}) - f(S^{(i-1)} \cup \{a\})\right| < \mathrm{rad}\right\} \middle| S^{(i-1)} = S\right] && \text{(definition of } \mathcal{E}_i) \\ &= \prod_{a \in \Omega \setminus S^{(i-1)}} \mathbb{P}\left[\left\{\left|\bar{f}(S^{(i-1)} \cup \{a\}) - f(S^{(i-1)} \cup \{a\})\right| < \mathrm{rad}\right\} \middle| S^{(i-1)} = S\right] \\ & && \text{(rewards are independent conditioned on actions)} \\ &\geq \left(1 - \frac{2}{T^4}\right)^{|\Omega \setminus S^{(i-1)}|} && \text{(using (6))} \\ &= \left(1 - \frac{2}{T^4}\right)^{n-i+1} \\ &\geq \left(1 - \frac{2}{T^4}\right)^{n}. && (7)\end{aligned}$$

Using (5) and (7), we are now ready to lower bound the probability of a clean event.

$$\begin{aligned}\mathbb{P}(\mathcal{E}) &= \mathbb{P}(\mathcal{E}_1 \cap \cdots \cap \mathcal{E}_k) \\ &= \prod_{i=1}^{k} \mathbb{P}(\mathcal{E}_i | \mathcal{E}_1, \ldots, \mathcal{E}_{i-1}) \\ &= \prod_{i=1}^{k} \sum_{S \in \left\{S' \,\middle|\, S' \subseteq \Omega,\ |S'| = i-1\right\}} \mathbb{P}(S^{(i-1)} = S | \mathcal{E}_1, \ldots, \mathcal{E}_{i-1}) \times \mathbb{P}(\mathcal{E}_i | S^{(i-1)} = S) && \text{(using (5))} \\ &\geq \prod_{i=1}^{k} \sum_{S \in \left\{S' \,\middle|\, S' \subseteq \Omega,\ |S'| = i-1\right\}} \mathbb{P}(S^{(i-1)} = S | \mathcal{E}_1, \ldots, \mathcal{E}_{i-1}) \times \left(1 - \frac{2}{T^4}\right)^{n} && \text{(using (7))} \\ &= \prod_{i=1}^{k} \left(1 - \frac{2}{T^4}\right)^{n} \sum_{S \in \left\{S' \,\middle|\, S' \subseteq \Omega,\ |S'| = i-1\right\}} \mathbb{P}(S^{(i-1)} = S | \mathcal{E}_1, \ldots, \mathcal{E}_{i-1}) \\ &= \prod_{i=1}^{k} \left(1 - \frac{2}{T^4}\right)^{n} \\ &= \left(1 - \frac{2}{T^4}\right)^{nk} \\ &\geq 1 - \frac{2nk}{T^4}. && \text{(Bernoulli's inequality)}\end{aligned}$$

This concludes the proof for Lemma 1.2. □

In Lemma 1.2, we showed that the clean event $\mathcal{E}$ will happen with high probability. Next, we present a lemma showing that the marginal gain of the action selected at the end of any exploitation phase is large under the condition that the clean event $\mathcal{E}$ happens.

Lemma 1.3 (Lemma 4.2 in Section 4)**.** *Under the clean event $\mathcal{E}$, for all $i \in \{1, \cdots, k\}$,*

$$f(S^{(i)}) - f(S^{(i-1)}) \geq \frac{1}{k}\left[f(S^*) - f(S^{(i-1)})\right] - 2\mathrm{rad}. \tag{8}$$

Proof. Recall that a_i, defined in (3), is the index of the arm that with $S^{(i-1)}$ forms the action with highest empirical mean at the end of phase i, i.e., $a_i = \arg\max_{a\in\mathcal{S}_i} \bar{f}(S^{(i-1)}\cup\{a\})$ and $S^{(i)} = S^{(i-1)}\cup\{a_i\}$. Let a_i^* denote the index of the arm that with $S^{(i-1)}$ forms the action with highest expected value, i.e, $a_i^* = \arg\max_{a\in\mathcal{S}_i} f(S^{(i-1)}\cup\{a\})$. For each $a \in \Omega\setminus S^{(i-1)}$, the event that the empirical mean $\bar{f}(S^{(i-1)}\cup\{a\})$ is concentrated within a radius of size rad around the expectated value can be written as

$$\begin{aligned}
& f(S^{(i-1)}\cup\{a\}) - \mathrm{rad} \leq \bar{f}(S^{(i-1)}\cup\{a\}) && \leq f(S^{(i-1)}\cup\{a\}) + \mathrm{rad} && \text{(concentration in } \mathcal{E}_i\text{)}\\
\Longleftrightarrow \quad & f(S^{(i-1)}\cup\{a\}) - 2\mathrm{rad} \leq \bar{f}(S^{(i-1)}\cup\{a\}) - \mathrm{rad} \leq f(S^{(i-1)}\cup\{a\}). && && (9)
\end{aligned}$$

We next lower bound the expected reward $f(S^{(i)})$ for the empirically best action in phase i, $S^{(i)} = \{a_i\}\cup S^{(i-1)}$. To do so, we apply (9) to two specific arms, the empirically best a_i and the statistically best a_i^*. We get

$$\begin{aligned}
f(S^{(i)}) &= f(S^{(i-1)}\cup\{a_i\}) && \text{(by design, } S^{(i)} \leftarrow \{a_i\}\cup S^{(i-1)}\text{)}\\
&\geq \bar{f}(S^{(i-1)}\cup\{a_i\}) - \mathrm{rad} && \text{(using (9))}\\
&\geq \bar{f}(S^{(i-1)}\cup\{a_i^*\}) - \mathrm{rad} && \text{(}a_i\text{ has the highest empirical mean)}\\
&\geq f(S^{(i-1)}\cup\{a_i^*\}) - 2\mathrm{rad}. && \text{(using (9))}
\end{aligned}$$

Subtracting $f(S^{(i-1)})$ on both side we have

$$f(S^{(i)}) - f(S^{(i-1)}) \geq f(S^{(i-1)}\cup\{a_i^*\}) - f(S^{(i-1)}) - 2\mathrm{rad}. \tag{10}$$

Recall from Section 2 that $S^* = \arg\max_{S:|S|\leq k} f(S)$ denotes the optimal solution in the offline problem. We will next show that the improvements in expectation of the chosen actions from one phase to the next are lower bounded by the gap between the optimal set S^* of cardinality k and the set $S^{(i)}$ chosen in the previous round.

$$\begin{aligned}
f(S^{(i)}) - f(S^{(i-1)}) &\geq f(S^{(i-1)}\cup\{a_i^*\}) - f(S^{(i-1)}) - 2\mathrm{rad} && \text{(copying (10))}\\
&= \max_{a\in\Omega\setminus S^{(i-1)}} f(S^{(i-1)}\cup\{a\}) - f(S^{(i-1)}) - 2\mathrm{rad} && \text{(by def.)}\\
&\geq \max_{a\in S^*\setminus S^{(i-1)}} f(S^{(i-1)}\cup\{a\}) - f(S^{(i-1)}) - 2\mathrm{rad} && \text{(restricted set)}\\
&\geq \frac{1}{|S^*\setminus S^{(i-1)}|}\sum_{a\in S^*\setminus S^{(i-1)}} f(S^{(i-1)}\cup\{a\}) - f(S^{(i-1)}) - 2\mathrm{rad} && \text{(max greater than average)}\\
&= \frac{1}{|S^*\setminus S^{(i-1)}|}\sum_{a\in S^*\setminus S^{(i-1)}} \left[f(S^{(i-1)}\cup\{a\}) - f(S^{(i-1)})\right] - 2\mathrm{rad} &&\\
&\geq \frac{1}{k}\sum_{a\in S^*\setminus S^{(i-1)}} \left[f(S^{(i-1)}\cup\{a\}) - f(S^{(i-1)})\right] - 2\mathrm{rad} && \text{(}S^*\text{ has cardinality }k\text{)}\\
&\geq \frac{1}{k}\left[f(S^*) - f(S^{(i-1)})\right] - 2\mathrm{rad}, && (11)
\end{aligned}$$

where (11) follows from a well known bound for submodular functions. □

Lemma 1.3 identifies a lower bound of the expected marginal gain $f(S^{(i)}) - f(S^{(i-1)})$ of the empirically best action $S^{(i)}$ at the end of phase i. As a corollary of Lemma 1.3, using properties of submodular set functions and unraveling the recursion induced by Lemma 1.3, we can lower bound the expected value of ETCG's chosen set $S^{(k)}$ of size k, which is used for exploitation in phase $k+1$.

Corollary 1.4 (Corollary 4.3 in Section 4)**.** *Under the clean event $\mathcal{E}$,*

$$f(S^{(k)}) \geq (1 - \frac{1}{e})f(S^*) - 2k\text{rad}. \tag{12}$$

Proof. We begin by unraveling the recursion induced by Lemma 1.3 and using properties of submodular set functions,

$$f(S^{(i)}) - f(S^{(i-1)}) \geq \frac{1}{k}\left[f(S^*) - f(S^{(i-1)})\right] - 2\text{rad}. \qquad \text{(copying (8))}$$

$$\begin{aligned} \iff \qquad f(S^{(i)}) &\geq \frac{1}{k}f(S^*) + (1 - \frac{1}{k})f(S^{(i-1)}) - 2\text{rad} && \text{(rearranging)} \\ &= \left[\frac{1}{k}f(S^*) - 2\text{rad}\right] + (1 - \frac{1}{k})f(S^{(i-1)}). && (13) \end{aligned}$$

Applying (13) recursively for $i = k$,

$$\begin{aligned} f(S^{(k)}) &\geq \left[\frac{1}{k}f(S^*) - 2\text{rad}\right] + (1 - \frac{1}{k})f(S^{(k-1)}) && \text{(using (13) for } i = k) \\ &\geq \left[\frac{1}{k}f(S^*) - 2\text{rad}\right] + (1 - \frac{1}{k})\left(\left[\frac{1}{k}f(S^*) - 2\text{rad}\right] + (1 - \frac{1}{k})f(S^{(k-2)})\right) && \text{(using (13) for } i = k-1) \\ &= \left[\frac{1}{k}f(S^*) - 2\text{rad}\right]\sum_{\ell=0}^{1}(1 - \frac{1}{k})^\ell + (1 - \frac{1}{k})^2 f(S^{(k-2)}) && \text{(rearranging)} \\ &\vdots && \text{(continue recursing until we get to } S^{(0)} = \emptyset;\ f(\emptyset) = 0) \\ &\geq \left[\frac{1}{k}f(S^*) - 2\text{rad}\right]\sum_{\ell=0}^{k-1}(1 - \frac{1}{k})^\ell && (14) \end{aligned}$$

Simplifying the geometric summation,

$$\begin{aligned} \sum_{\ell=0}^{k-1}(1 - \frac{1}{k})^\ell &= \frac{1 - (1 - \frac{1}{k})^k}{1 - (1 - \frac{1}{k})} \\ &= k\left(1 - (1 - \frac{1}{k})^k\right). \end{aligned}$$

Continuing with (14),

$$\begin{aligned} f(S^{(k)}) &\geq \left[\frac{1}{k}f(S^*) - 2\text{rad}\right] k\left(1 - (1 - \frac{1}{k})^k\right) \\ &= \left(1 - \left(1 - \frac{1}{k}\right)^k\right) f(S^*) - 2k\left(1 - (1 - \frac{1}{k})^k\right)\text{rad} \\ &\geq \left(1 - \left(1 - \frac{1}{k}\right)^k\right) f(S^*) - 2k\text{rad}. && \text{(simplifying with } (1 - \tfrac{1}{k})^k \leq 1) \end{aligned}$$

Using the well-known lower bound $\left(1 - \left(1 - \frac{1}{k}\right)^k\right) \geq 1 - \frac{1}{e}$, we get

$$f(S^{(k)}) \geq (1 - \frac{1}{e})f(S^*) - 2k\text{rad}.$$

Rearranging terms we have

$$(1-\frac{1}{e})f(S^*) - f(S^{(k)}) \leq 2k\text{rad}.$$

□

1.2 THEOREM 4.1 PROOF

Now we are ready to prove the main theorem, Theorem 4.1.

Case 1: clean event $\mathcal{E}$ happens

In the first case we analyse the expected regret under the condition that the clean event $\mathcal{E}$ happens. In this section, all expectations will be conditioned on $\mathcal{E}$, but to simplify notation we will write $\mathbb{E}[\cdot]$ instead of $\mathbb{E}[\cdot|\mathcal{E}]$.

First we can break up the expected $(1-\frac{1}{e})$-regret (2) conditioned on $\mathcal{E}$ into two parts, one for the first k phases, and the second for the exploitation phase. Also recall that $f_t(S_t)$ is the random reward for taking action S_t, which itself is random, depending on empirical means of actions in earlier phases.

$$\begin{aligned}
\mathbb{E}[\mathcal{R}(T)] &= (1-\frac{1}{e})Tf(S^*) - \sum_{t=1}^{T}\mathbb{E}[f_t(S_t)] && \text{(using the definition (2))} \\
&= (1-\frac{1}{e})Tf(S^*) - \sum_{t=1}^{T}\mathbb{E}[\mathbb{E}[f_t(S_t)|S_t]] && \text{(law of total expectation)} \\
&= (1-\frac{1}{e})Tf(S^*) - \sum_{t=1}^{T}\mathbb{E}[f(S_t)] && \text{($f(\cdot)$ defined as expected reward)} \\
&= \sum_{t=1}^{T}\left((1-\frac{1}{e})f(S^*) - \mathbb{E}[f(S_t)]\right) && \text{(rearranging)} \\
&= \underbrace{\sum_{i=1}^{k}\sum_{t=T_{i-1}+1}^{T_i}\left((1-\frac{1}{e})f(S^*) - \mathbb{E}[f(S_t)]\right)}_{\text{First } k \text{ phases}} + \underbrace{\sum_{t=T_k+1}^{T}\left((1-\frac{1}{e})f(S^*) - \mathbb{E}[f(S_t)]\right)}_{\text{Exploitation phase}} \\
&= \sum_{i=1}^{k}\sum_{t=T_{i-1}+1}^{T_i}\left((1-\frac{1}{e})f(S^*) - \mathbb{E}[f(S_t)]\right) + \sum_{t=T_k+1}^{T}\left((1-\frac{1}{e})f(S^*) - \mathbb{E}[f(S^{(k)})]\right). && (15)
\end{aligned}$$

Recall that in phase i, each of the $n-i+1$ actions in $\mathcal{S}_i$ is played exactly m times, meaning $T_i - T_{i-1} = m(n-i+1)$. Since all actions played in phase i include the set $S^{(i-1)}$ (the empirically best set played in phase $i-1$), in notation $S^{(i-1)} \subset S_t$ for $t \in \{T_{i-1}+1, \cdots, T_i\}$, by monotonicity of the expected reward function f, we have $f(S^{(i-1)}) \leq f(S_t)$, for $t \in \{T_{i-1}+1, \cdots, T_i\}$. Thus, we can simplify the inner summation in the first term of (15) as

$$\begin{aligned}
\sum_{t=T_{i-1}+1}^{T_i}\left((1-\frac{1}{e})f(S^*) - \mathbb{E}[f(S_t)]\right) &\leq \sum_{t=T_{i-1}+1}^{T_i}\left((1-\frac{1}{e})f(S^*) - \mathbb{E}[f(S^{(i-1)})]\right) \\
& \qquad\qquad \text{(monotonicity: } f(S^{(i-1)}) \leq f(S_t)) \\
&= m(n-i+1)\left((1-\frac{1}{e})f(S^*) - \mathbb{E}[f(S^{(i-1)})]\right). && (16)
\end{aligned}$$

Plugging (16) back into (15),

$$\begin{aligned}\mathbb{E}[\mathcal{R}(T)] &\leq \sum_{i=1}^{k} m(n-i+1)\left((1-\frac{1}{e})f(S^*) - \mathbb{E}[f(S^{(i-1)})]\right) + \sum_{t=T_k+1}^{T}\left((1-\frac{1}{e})f(S^*) - \mathbb{E}[f(S^{(k)})]\right) \\ &\leq mn\sum_{i=1}^{k}\left((1-\frac{1}{e})f(S^*) - \mathbb{E}[f(S^{(i-1)})]\right) + \sum_{t=T_k+1}^{T}\left((1-\frac{1}{e})f(S^*) - \mathbb{E}[f(S^{(k)})]\right). \qquad (17)\end{aligned}$$

Now we upper bound the two terms above using Corollary 1.4.

Since for $i \in \{2, \cdots, k\}$, $S^{(i-1)}$'s are random variables, we can take the expectation of (8) (conditioned on event $\mathcal{E}$), yielding

$$\begin{aligned} & \mathbb{E}[f(S^{(i)})] - \mathbb{E}[f(S^{(i-1)})] \geq \frac{1}{k}\left[f(S^*) - \mathbb{E}[f(S^{(i-1)})]\right] - 2\text{rad}, \qquad (18) \\ \iff \qquad & f(S^*) - \mathbb{E}[f(S^{(i-1)})] \leq k(\mathbb{E}[f(S^{(i)})] - \mathbb{E}[f(S^{(i-1)})] + 2\text{rad}). \qquad (19)\end{aligned}$$

and of (12), yielding

$$(1-\frac{1}{e})f(S^*) - \mathbb{E}[f(S^{(k)})] \leq 2k\text{rad}. \qquad (20)$$

Apply (19) and (20) to the first and second terms in (17) respectively yields

$$\begin{aligned}\mathbb{E}[\mathcal{R}(T)] &\leq mn\sum_{i=1}^{k}\left((1-\frac{1}{e})f(S^*) - \mathbb{E}[f(S^{(i-1)})]\right) + \sum_{t=T_k+1}^{T}\left((1-\frac{1}{e})f(S^*) - \mathbb{E}[f(S^{(k)})]\right) && \text{(copying (17))} \\ &\leq mn\sum_{i=1}^{k}\left(f(S^*) - \mathbb{E}[f(S^{(i-1)})]\right) + \sum_{t=T_k+1}^{T}\left((1-\frac{1}{e})f(S^*) - \mathbb{E}[f(S^{(k)})]\right) && \text{(using } 1-\tfrac{1}{e} \leq 1 \text{ in first sum)} \\ &\leq mnk\sum_{i=1}^{k}\left(\mathbb{E}[f(S^{(i)})] - \mathbb{E}[f(S^{(i-1)})] + 2\text{rad}\right) + \sum_{t=T_k+1}^{T}(2k\text{rad}) && \text{(using (19) and (20))} \\ &= mnk\left(\mathbb{E}[f(S^{(k)})] - \mathbb{E}[f(S^{(0)})] + 2k\text{rad}\right) + \sum_{t=T_k+1}^{T}(2k\text{rad}) && \text{(telescoping sum)} \\ &\leq mnk\left(\mathbb{E}[f(S^{(k)})] + 2k\text{rad}\right) + 2kT\text{rad} && (f(S^{(0)}) = 0) \\ &\leq mnk\left(1 + 2k\text{rad}\right) + 2kT\text{rad}. && \text{(rewards are bounded in } [0,1])\end{aligned}$$

Plugging in the definition of $\text{rad} = \sqrt{2\log(T)/m}$ and using the bound $\sqrt{2\log(T)/m} < \sqrt{2\log(T)}$ to simplify the formula, we have

$$\begin{aligned}\mathbb{E}[\mathcal{R}(T)] &\leq mnk\left(1 + 2k\sqrt{2\log(T)/m}\right) + 2kT\sqrt{2\log(T)/m} \\ &\leq mnk\left(1 + 2k\sqrt{2\log(T)}\right) + 2kT\sqrt{2\log(T)/m}. \qquad (21)\end{aligned}$$

We want to optimize m, the number of times actions are played. Denoting the regret bound (21) as a function of m

$$g(m) = mnk\left(1 + 2k\sqrt{2\log(T)}\right) + 2kT\sqrt{2\log(T)/m}, \qquad (22)$$

then

$$g'(m) = nk\left(1 + 2k\sqrt{2\log(T)}\right) - kT\sqrt{2\log(T)}m^{-3/2}. \qquad (23)$$

Setting $g'(m) = 0$ and solving for m,

$$m^* = \left(\frac{T\sqrt{2\log(T)}}{n + 2nk\sqrt{2\log(T)}} \right)^{2/3}. \tag{24}$$

We next check the second derivative,

$$g''(m) = \frac{3}{2} kT\sqrt{2\log(T)} m^{-5/2}. \tag{25}$$

For positive values of m, $g''(m) > 0$, thus $g(m)$ reaches a minima at (24).

Since m is the number of times actions are played, we (trivially) need $m \geq 1$ and m to be an integer. We choose

$$m^\dagger = \left\lceil \left(\frac{T\sqrt{2\log(T)}}{n + 2nk\sqrt{2\log(T)}} \right)^{2/3} \right\rceil. \tag{26}$$

Since from (25) we have that $g''(m) > 0$ for positive m, $g(m^*) \leq g(m^\dagger)$.

For $T \geq n(k+1)$, we have

$$\begin{aligned} m^* = \left(\frac{T\sqrt{2\log(T)}}{n + 2nk\sqrt{2\log(T)}} \right)^{2/3} &= \left(\frac{T}{\frac{n}{\sqrt{2\log(T)}} + 2nk} \right)^{2/3} \\ &\geq \left(\frac{n(k+1)}{\frac{n}{\sqrt{2\log(n(k+1))}} + 2nk} \right)^{2/3} \\ &= \left(\frac{k+1}{\frac{1}{\sqrt{2\log(n(k+1))}} + 2k} \right)^{2/3} \\ &\geq \left(\frac{k+1}{2k+1} \right)^{2/3} \\ &\geq \left(\frac{1}{2} \right)^{2/3} \\ &> \frac{1}{2}. \end{aligned} \tag{27}$$

Plugging (26) back in to (21),

$$\begin{aligned}
\mathbb{E}[\mathcal{R}(T)] &\leq m^{\dagger}nk\left(1+2k\sqrt{2\log(T)}\right)+2kT\sqrt{2\log(T)/m^{\dagger}} && \text{((21) with } m^{\dagger} \text{ samples for each action)}\\
&= \lceil m^*\rceil nk\left(1+2k\sqrt{2\log(T)}\right)+2kT\sqrt{2\log(T)/\lceil m^*\rceil}\\
&\leq \lceil m^*\rceil nk\left(1+2k\sqrt{2\log(T)}\right)+2kT\sqrt{2\log(T)/m^*} && \text{(Since } \lceil m^*\rceil \geq m^*\text{)}\\
&\leq 2m^*nk\left(1+2k\sqrt{2\log(T)}\right)+2kT\sqrt{2\log(T)/m^*} && \text{(Since } m^*\geq 1/2,\ \lceil m^*\rceil \leq 2m^*\text{)}\\
&= 2\left(\frac{T\sqrt{2\log(T)}}{n+2nk\sqrt{2\log(T)}}\right)^{2/3} nk(1+2k\sqrt{2\log(T)})+2kT\sqrt{2\log(T)}\left(\frac{n+2nk\sqrt{2\log(T)}}{T\sqrt{2\log(T)}}\right)^{1/3}\\
& && \text{(using (24))}\\
&= \frac{2(T\sqrt{2\log(T)})^{2/3}}{n^{2/3}(1+2k\sqrt{2\log(T)})^{2/3}}nk(1+2k\sqrt{2\log(T)})+2kT\sqrt{2\log(T)}\frac{n^{1/3}(1+2k\sqrt{2\log(T)})^{1/3}}{(T\sqrt{2\log(T)})^{1/3}}\\
& && \text{(rearranging)}\\
&= 2(T\sqrt{2\log(T)})^{2/3}n^{1/3}k\left(1+2k\sqrt{2\log(T)}\right)^{1/3}+2k(T\sqrt{2\log(T)})^{2/3}n^{1/3}(1+2k\sqrt{2\log(T)})^{1/3}\\
& && \text{(cancelling common terms)}\\
&= 4n^{\frac{1}{3}}k(T\sqrt{2\log(T)})^{\frac{2}{3}}(1+2k\sqrt{2\log(T)})^{\frac{1}{3}} && (28)\\
&= \mathcal{O}(n^{\frac{1}{3}}k^{\frac{4}{3}}T^{\frac{2}{3}}\log(T)^{\frac{1}{2}}).
\end{aligned}$$

where (28) follows by factoring. In conclusion, the expected $(1-1/e)$ regret (2) is upper bounded by (28) if the clean event $\mathcal{E}$ happens.

Case 2: clean event $\mathcal{E}$ does not happen

We next derive an upper bound for the expected $(1-1/e)$ regret (2) for case that the event $\mathcal{E}$ does not happen. By Lemma 1.2,

$$\mathbb{P}(\bar{\mathcal{E}}) = 1-\mathbb{P}(\mathcal{E}) \leq \frac{2nk}{T^4}.$$

Since the reward function $f_t(\cdot)$ is upper bounded by 1, the expected $(1-1/e)$ regret (2) incurred under $\bar{\mathcal{E}}$ for a horizon of T is at most T,

$$\mathbb{E}[\mathcal{R}(T)|\bar{\mathcal{E}}] \leq T. \tag{29}$$

Putting it all together

Combining Cases 1 and 2 we have,

$$\begin{aligned}
\mathbb{E}[\mathcal{R}(T)] &= \mathbb{E}[\mathcal{R}(T)|\mathcal{E}]\cdot\mathbb{P}(\mathcal{E})+\mathbb{E}[\mathcal{R}(T)|\bar{\mathcal{E}}]\cdot\mathbb{P}(\bar{\mathcal{E}}) && \text{(Law of total expectation)}\\
&\leq \left[4n^{\frac{1}{3}}k(T\sqrt{2\log(T)})^{\frac{2}{3}}(1+2k\sqrt{2\log(T)})^{\frac{1}{3}}\right]\cdot 1+T\cdot 2nkT^{-4} && \text{(using (28), Lemma 1.2, and (29))}\\
&= \mathcal{O}(n^{\frac{1}{3}}k^{\frac{4}{3}}T^{\frac{2}{3}}\log(T)^{\frac{1}{2}}).
\end{aligned}$$

This concludes the proof of Theorem 4.1.

2 ALGORITHM OG$^{\mathrm{o}}$

In this section we describe implementation details and parameter selection for OG$^{\mathrm{o}}$ algorithm Streeter and Golovin [2008]. The choice of exploration probability is given by the original paper:$\gamma = n^{1/3}k\left(\frac{\log(n)}{T}\right)^{1/3}$. ϵ is the learning rate for Randomized Weighted Majority (WMR) expert algorithm Arora et al. [2012]. It is chosen by setting the derivative of regret upper bound to zero, which is $\epsilon = \sqrt{\frac{\log(n)}{T_e}}$, where T_e is the time spent on updating expert e. Since it explores with probability γ, and there are k expert algorithms, we have $T_e \approx \frac{\gamma T}{k}$. Thus we pick $\epsilon = \sqrt{\frac{k\log(n)}{\gamma T}}$. In experiments, there are many cases the chosen γ is large or even larger than 1, so we cap the probability of exploring γ by 1/2 to avoid exploring too much. Algorithm 2 shows the pseudo code for implementation details of this algorithm.

Algorithm 2 Online Greedy for Opaque Feedback Model (OG$^{\text{o}}$)

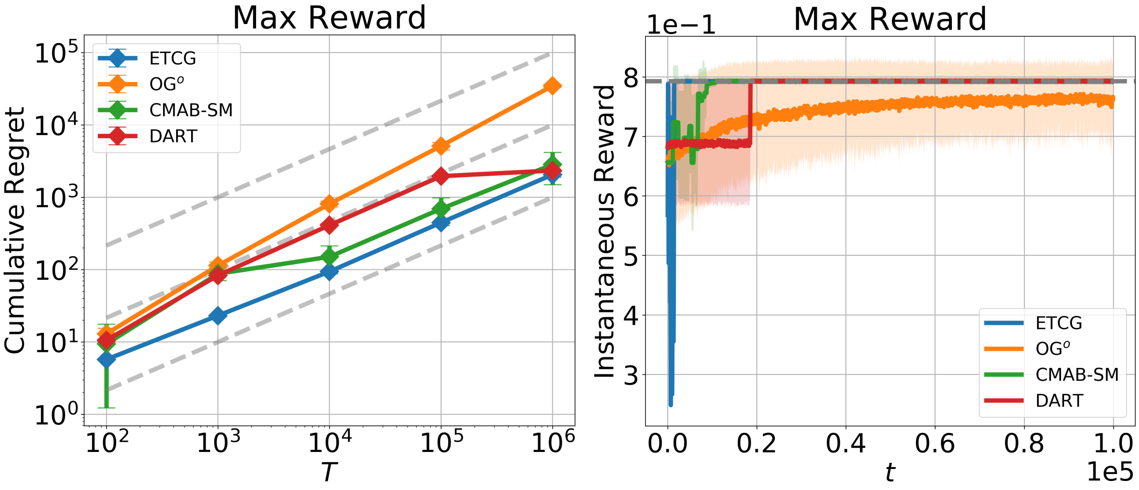
```
Input: set of base arms $\Omega$, horizon $T$, cardinality constraint $k$
Initialize $n \leftarrow |\Omega|, \gamma \leftarrow n^{1/3}k\left(\frac{\log(n)}{T}\right)^{1/3}, \epsilon \leftarrow \sqrt{\frac{k\log(n)}{\gamma T}}$
Initialize $\boldsymbol{\omega}_1 \leftarrow$ ones$(k, n)$
for $t \in [1, \cdots, T]$ do
    $S_t \leftarrow \emptyset$
    $l \leftarrow$ zeros$(k, n)$                                              ▷ loss
    Randomly sample a value $\xi \sim$ Uniform$([0, 1])$
    if $\xi \leq \gamma$ then                                      ▷ Exploration with probability $\gamma$
        $e \sim$ Uniform$(\{1, \cdots, k\})$
        for $i \in [1, \cdots, e-1]$ do                           ▷ For experts before $e$, exploit
            Select an arm $a$ with probability $\frac{\boldsymbol{\omega}_t[i,a]}{\sum \boldsymbol{\omega}_t[i,:]}$, re-sample if $a \in S_t$
            $S_t \leftarrow S_t \cup \{a\}$
        end for
        $a \sim$ Uniform$(\{1, \cdots, n\} \backslash S_t)$                           ▷ For expert $e$, explore
        $S_t \leftarrow S_t \cup \{a\}$
        Play action $S_t$, observe $f_t(S_t)$
        Update $l[i, j] \leftarrow f_t(S_t)$ for all $i = e$ and $j \neq a$      ▷ Feed back $f_t(S_t)$ to expert $e$ associated with action $a$
        Update $\boldsymbol{\omega}_{t+1}[i, j] \leftarrow \boldsymbol{\omega}_t[i, j]\exp(-\epsilon l[i, j])$ for all pairs of $i$ and $j$
    else                                                ▷ Exploitation with probability $1 - \gamma$
        for $i \in [1, \cdots, k]$ do                             ▷ For experts before $e$, exploit
            Select arm $a$ with probability $\frac{\boldsymbol{\omega}_t[i,a]}{\sum \boldsymbol{\omega}_t[i,:]}$, re-sample if $a \in S_t$
            $S_t \leftarrow S_t \cup \{a\}$
        end for
        Play action $S_t$, observe $f_t(S_t)$
        $\boldsymbol{\omega}_{t+1}[i, j] \leftarrow \boldsymbol{\omega}_t[i, j]$          ▷ Since feeding back 0 to all expert-action payoffs, loss is 0, no update
    end if
end for
```

3 MORE EXPERIMENTS

3.1 MAX FUNCTION

We also conduct experiments with synthetic data on max functions: $f(S) = \max_{a \in S} f(\{a\})$. Similar with the setup in Section 5.2, We use $n = 20$ base arms and cardinality constraint $k = 4$. Again, we generate individual arm rewards $\{f(\{a\})\}_{a \in \Omega}$ randomly $f(\{a\}) \overset{i.i.d.}{\sim} \mathcal{U}([0.1, 0.9])$ and add noise when sampling. The noise follows a truncated normal distribution with mean 0 and standard deviation 0.1 within interval $[-0.1, 0.1]$. The results are shown in Figure 1.

We can see from Figure 1a, ETCG outperforms all other baseline methods evaluated up to $T = 10^6$, but DART seems to be able to surpass ETCG for larger T. The reason is that max reward function bounded in $[0, 1]$ satisfies the assumptions of DART, so DART's $\mathcal{O}(T^{1/2})$ regret bound holds. Thus, we expect DART to eventually outperform ETCG for max reward functions. Notably, despite DART's asymptotic advantage for $\max$ function, ETCG does better than DART for all but very large horizons (namely T=1,000,000). We argue it is unrealistic for any application to be stationary (assumed by DART) over such a long horizon.



(a)

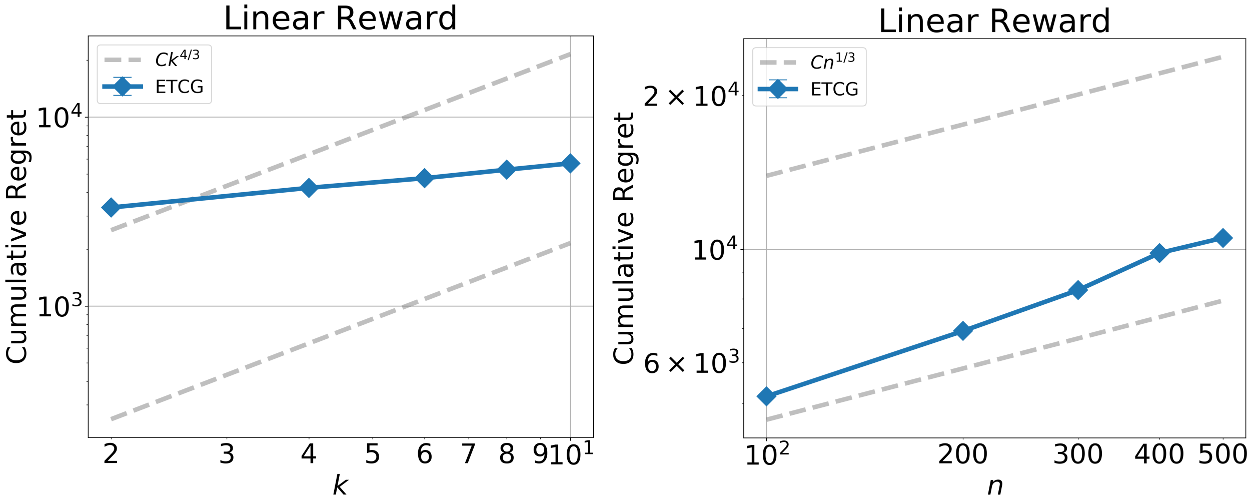
(b)

Figure 1: (a) shows results for cumulative regret as a function of time horizon T. (b) shows the moving average plot with window size 100 of instantaneous reward as a function of t. The gray dashed lines in (a) represent $y = aT^{2/3}$ for various values of a. The gray dashed line in (b) represents the value of the optimal solution.

3.2 DENPENDENCE ON n AND k

We also empirically plot the regret as a function of n and k to see if the dependence on n and k is "correct" for linear functions.

The results are shown in Figure 2. From the figures we can see that for linear rewards, $\mathcal{O}(n^{1/3})$ appears tight and $\mathcal{O}(k^{4/3})$ appears loose (the estimated exponent is closer to $O(k^{1/3})$). We will leave it as an open question on whether there exists an algorithm that has a better guarantee with respect to k.



(a)

(b)

Figure 2: (a) shows results for cumulative regret as a function of cardinality constraint k. (b) shows results for cumulative regret as a function of number of base arms n. The gray dashed lines in (a) represent $y = aT^{4/3}$ for various values of a. The gray dashed lines in (b) represent $y = CT^{1/3}$ for various values of a.

References

Sanjeev Arora, Elad Hazan, and Satyen Kale. The multiplicative weights update method: a meta-algorithm and applications. *Theory Comput.*, 8:121–164, 2012.

George L Nemhauser, Laurence A Wolsey, and Marshall L Fisher. An analysis of approximations for maximizing submodular set functions—i. *Mathematical programming*, 14(1):265–294, 1978.

Aleksandrs Slivkins. Introduction to multi-armed bandits. *Foundations and Trends® in Machine Learning*, 12(1-2):1–286, 2019. ISSN 1935-8237.

Matthew Streeter and Daniel Golovin. An online algorithm for maximizing submodular functions. In *Proceedings of the 21st International Conference on Neural Information Processing Systems*, NIPS'08, page 1577–1584, Red Hook, NY, USA, 2008. Curran Associates Inc.