Logarithmic Regret for Unconstrained Submodular Maximization Stochastic Bandit

Julien Zhou, Pierre Gaillard, Thibaud Rahier, Julyan Arbel

Logarithmic Regret for Unconstrained Submodular Maximization Stochastic Bandit

Julien Zhou JULIEN.ZHOU@INRIA.FR
Criteo AI Lab, Paris, France
Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France
Pierre Gaillard PIERRE.GAILLARD@INRIA.FR
Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France
Thibaud Rahier T.RAHIER@CRITEO.COM
Criteo AI Lab, Paris, France
Julyan Arbel JULYAN.ARBEL@INRIA.FR
Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France

Abstract

We address the *online unconstrained submodular maximization problem* (Online USM), in a setting with *stochastic bandit feedback*. In this framework, a decision-maker receives noisy rewards from a nonmonotone submodular function, taking values in a known bounded interval. This paper proposes *Double-Greedy - Explore-then-Commit* (`DG-ETC`), adapting the Double-Greedy approach from the offline and online full-information settings. `DG-ETC` satisfies a $O(d\log(dT))$ problem-dependent upper bound for the $1/2$-approximate pseudo-regret, as well as a $O(dT^{2/3}\log(dT)^{1/3})$ problem-free one at the same time, outperforming existing approaches. To that end, we introduce a notion of hardness for submodular functions, characterizing how difficult it is to maximize them with this type of strategy.

Keywords: Unconstrained submodular maximization; stochastic bandits; logarithmic regret; sampling complexity.

1. Introduction

1.1. Context and problem formulation

Several real-world settings can cast as combinatorial optimization problems over a finite set. Without some assumptions on the utility function to be maximized and/or the constraints to be satisfied, such problems are not solvable in polynomial time. In practice, different types of assumptions and constraints can be introduced to make these problems manageable, even approximately. One can, for example, assume the utility to be linear. However, even this already strong assumption can be helpless in making the problem easier, depending on the constraints.

In this paper, we focus on the cases where the maximized function is submodular, meaning that it satisfies a "diminishing marginal gains" property. We consider the unconstrained setting, where the whole combinatorial super-set is available and we do not assume the utility to be monotonous (if it is, the solution is straightforward). We also place ourselves in a stochastic (combinatorial) bandit setting, where a *decision-maker / player* chooses different sets along sequential rounds, and receives noisy rewards. In this framework, one classically needs to balance exploration and exploitation, but also to manage the combinatorial complexity of the problem at the same time. In particular, a good strategy should efficiently leverage the underlying structure of the reward – submodularity in this case – and estimate/exploit relevant quantities.

Problem formulation and assumptions. We consider a finite set of items $\mathcal{D}$ where $|\mathcal{D}| = d \in \mathbb{N}^*$. The player has access to actions from the whole superset $\mathcal{P}(\mathcal{D})$ and plays for an horizon of $T \in \mathbb{N}^*$ rounds. We assume that the player receives noisy rewards from a *nonmonotonous submodular* set-function $f : \mathcal{P}(\mathcal{D}) \to [0, c]$ with $c > 0$. At each round $t \in [T] = \{1, \dots, T\}$, the player chooses an action $A_t \in \mathcal{P}(\mathcal{D})$ and receives

$$Z_t = f(A_t) + \eta_t \,, \tag{1}$$

where η_t is a random variable. Let $\sigma > 0$ (known), we assume that η_t is σ^2-sub-Gaussian conditionally to $\mathcal{F}_{t-1} = \sigma(\eta_1, \dots, \eta_{t-1})$ and independent to the possibly random process generating the actions $(A_s)_{s \leq t}$.

The algorithms that we study in this paper all consider the items sequentially. For convenience, we identify $\mathcal{D}$ with $[d]$ and assume an arbitrary ordering, but a player with prior knowledge could try to optimize over permutations.

$1/2$-Approximate pseudo-regret minimization. The objective of the player is to maximize its cumulative rewards over the T rounds. As it is common in bandit literature, we instead look at a pseudo-regret, neglecting the contribution of the noise $(\eta_t)_{t \in [T]}$. Besides, rather than looking at the exact pseudo-regret, we minimize an $1/2$-approximation defined as

$$R_T = \sum_{t \in [T]} \left[\frac{1}{2} f(A^*) - f(A_t) \right], \tag{2}$$

where $A^* \in \arg\max_{A \subseteq \mathcal{D}} \{ f(A) \}$. Considering approximate regrets is usual in settings where we have access to an *oracle* solving the offline optimization approximately (Chen et al., 2013). In particular, the $1/2$ factor here come from impossibility of solving the offline unconstrained submodular maximization problem (USM), with a competitive ratio better than $1/2$, using a polynomial number of calls (Feige et al., 2011). In the following, except if we specify it explicitly, the expressions *pseudo-regret* or even just *regret* refer to the $1/2$-*approximate pseudo-regret*.

1.2. Related works

Combinatorial bandits. The recent monograph by Lattimore and Szepesvári (2020) makes an extensive study of bandit problems. We are more particularly interested in settings where the action space is combinatorial and too big to be explored in its entirety. Chen et al. (2013) introduced a semi-bandit stochastic framework with approximate regrets and a monotone aggregation function having bounded smoothness. When the aggregation is linear, the leading factors in the regrets have been refined in several subsequent works (Kveton et al., 2015; Degenne and Perchet, 2016; Perrault et al., 2020; Zhou et al., 2024). A matching adversarial semi-bandit setting has also been explored (Ito, 2021; Neu and Valko, 2014). Besides, when the feedback is "only" full-bandit, and the aggregation remains linear, one could see the problem as a linear bandit and use the corresponding methods, as long as the offline problem can be solved (Abbasi-Yadkori et al., 2011; Bubeck et al., 2012). Considering a nonlinear aggregation function with full-bandit feedback remains however challenging without further assumptions on the reward (Han et al., 2021).

Unconstrained submodular maximization (USM). Submodularity has applications in various fields, including economics, game theory and combinatorial optimization. As it shares properties

similar to both convexity and concavity in continuous optimization (Lovász, 1983), both viewpoints are of interest. The monograph by Bach (2013) details various cases where submodular set-functions appear and highlights the parallels between submodular minimization and convex optimization. While submodular minimization can be solved exactly in polynomial time, the maximization is more difficult, and can in general be solved only approximately (Feige et al., 2011). A $(1-1/e)$-approximation is possible in the cardinally-constrained monotone case (Nemhauser et al., 1978), but the unconstrained nonmonotone setting can only be solved up to a $1/2$-approximation (Feige et al., 2011). Buchbinder et al. (2012) particularly shows that a double-greedy approach reaches this ratio using a linear number of oracle calls.

Online USM with partial feedback. While Feige et al. (2011) proves that a $1/2$-approximation cannot be improved with a polynomial number of oracle calls in the offline setting, Buchbinder et al. (2012) provides a linear-time approach reaching this ratio, closing the gap between upper and lower bounds. Later on, Roughgarden and Wang (2018) introduces an online full-information adversarial framework and provides an algorithm satisfying a $O(d\sqrt{T})$ upper bound for the expected $1/2$-approximate regret. Harvey et al. (2020) manages to gain a $\sqrt{d}$ factor on this bound by using tools related to online dual averaging and Blackwell approachability. However, the bandit setting where the player has considerably less feedback has been less studied. Fourati et al. (2023) considers a stochastic bandit setting, and proposes an Explore-then-Commit type algorithm satisfying a $O(dT^{2/3}\log(T)^{1/3})$ upper bound. However, Niazadeh et al. (2021) claims a similar $O(dT^{2/3})$ in an adversarial bandit setting. As the latter framework seems significantly more difficult, one may reasonably wonder if better guarantees can be satisfied in the stochastic setting. We answer this question positively and propose an algorithm satisfying both logarithmic problem-dependent and $O(d(T\log(dT))^{2/3})$ problem-free upper bounds.

1.3. Contributions

We propose a novel algorithm *Double-Greedy - Explore-then-commit* (`DG-ETC`) for the online unconstrained submodular maximization problem (Online USM), with stochastic bandit feedback (Section 3). We introduce a new notion of *hardness* for this problem (Section 4.1), and prove that `DG-ETC` satisfies both a logarithmic problem-dependent (hardness-dependent) upper bound for the $1/2$-approximate pseudo-regret, as well as a worst-case $O(dT^{2/3}\log(dT)^{1/3})$ (Sections 4.2 and 5). Those bounds are satisfied both with high-probability and in expectation (Theorem 2) and leverage the differences between the stochastic setting (Fourati et al., 2023) and the adversarial one (Niazadeh et al., 2021). Asymptotically, `DG-ETC` allocates a logarithmic, hardness-dependent, number of rounds to the design of a strategy that compensates the randomness errors with per-round negative losses (therefore, with gains). `DG-ETC` actually leverages the looseness of the $1/2$-approximation ratio in non-adversarial cases, and we argue that this kind of strategy could also be applied to other settings involving approximations.

2. Preliminary

In this section, we introduce submodularity, and remind the spirit of the Double-Greedy algorithm (Buchbinder et al., 2012) on which our `DG-ETC` is based.

2.1. Submodularity

Submodularity is a "diminishing marginal gains" property. It is formally defined as follows.

Definition 1 (Submodularity) *Let $\mathcal{D}$ be a finite set and $c > 0$. A set-function $f : \mathcal{P}(\mathcal{D}) \to [0, c]$ is said to be (bounded)* submodular *if, equivalently,*

- *For all $A \subseteq B \subseteq \mathcal{D}$ and $i \in \mathcal{D}$,* $\quad f(B \cup \{i\}) - f(B) \leq f(A \cup \{i\}) - f(A)$;
- *For all $(A,\ B) \in \mathcal{P}(\mathcal{D}) \times \mathcal{P}(\mathcal{D})$,* $\quad f(A \cup B) + f(A \cap B) \leq f(A) + f(B)$.

Besides, f is said to be monotone *if for all $A \subseteq B \subseteq \mathcal{D}$, $f(A) \leq f(B)$. Otherwise, we say that f is* nonmonotone.

2.2. Double-Greedy for USM

Understanding the Double-Greedy Algorithm (DG, Algorithm 1) from Buchbinder et al. (2012) is crucial for the rest of the paper.

When maximizing a nonmonotone submodular function f, DG works in d steps (one per item) and considers the items sequentially.

It first initializes a pair of sets $X_0 = \emptyset$ and $Y_0 = \mathcal{D}$ as the empty set and the full set respectively, and then modifies them sequentially.

At each step $i \in [d]$, DG looks at the "marginal gains" α_i and β_i respectively corresponding to adding item i to X_{i-1} or removing it from Y_{i-1} [Line 4-5]. It makes the final decision of either adding or removing it by sampling a Bernoulli random variable with parameter p_i, defined from the positive part of the marginal gains [Line 6-7]. After the d-th and last step, DG returns the set X_d, which is identical to Y_d by construction [Line 14].

Overall, DG requires $4d$ calls to f and satisfies the following guarantee.

Algorithm 1 Double-Greedy (DG from Buchbinder et al., 2012)

1: **Inputs:** $\mathcal{D}$
2: $(X_0,\ Y_0) \leftarrow (\emptyset,\ \mathcal{D})$
3: **for** $i = 1, \dots d$ **do**
4: $\quad \alpha_i \leftarrow f(X_{i-1} \cup \{i\}) - f(X_{i-1})$
5: $\quad \beta_i \leftarrow f(Y_{i-1} \setminus \{i\}) - f(Y_{i-1})$
6: $\quad p_i \leftarrow \frac{\max\{\alpha_i, 0\}}{\max\{\alpha_i, 0\} + \max\{\beta_i, 0\}}$
7: $\quad K_i \sim \mathcal{B}(p_i)$
8: $\quad$ **if** K_i **then**
9: $\quad\quad (X_i,\ Y_i) \leftarrow (X_{i-1} \cup \{i\},\ Y_{i-1})$
10: $\quad$ **else**
11: $\quad\quad (X_i,\ Y_i) \leftarrow (X_{i-1},\ Y_{i-1} \setminus \{i\})$
12: $\quad$ **end if**
13: **end for**
14: **Return:** $X_d \subseteq \mathcal{D}$

Theorem 1 (Buchbinder et al. (2012), Theorem I.2.) *Let $\mathcal{D}$ be a finite set. Algorithm DG returns a set S such that $\mathbb{E}\big[f(S)\big] \geq \frac{1}{2} f(A^*)$.*

The result being in expectation, one can repeatedly run DG to obtain an acceptable set with a high enough probability. In particular, we prove the following Proposition in Appendix B.

Proposition 1 *Let $\mathcal{D}$ be a finite set, $\delta > 0$ and $T \in \mathbb{N}^*$ such that $T > 2\log(1/\delta)$. If $(S_i)_{i \in [T]}$ is the sequence of sets obtained by running independently T times DG Algorithm, then*

$$\max_{i \in [T]} f(S_i) > \Big(\frac{1}{2} - \frac{\log(1/\delta)}{T}\Big) f(A^*)\,, \qquad \textit{w.p.} \quad 1 - \delta\,.$$

Stochastic bandit setting. In our setting, using DG directly is not possible as we do not have access to the marginal gains α_i and β_i but only to noisy estimates. To overcome this difficulty, Fourati et al. (2023) propose the *Randomized Greedy Learning* (RGL) algorithm, an *Explore-then-Commit* strategy satisfying a $O(dT^{2/3}\log(T)^{1/3})$ expected regret upper bound. Similarly to DG, RGL works in d steps, one per item, each lasting $T^{2/3}\log(T)^{1/3}$ rounds. During the i-th step, RGL estimates the coefficients α_i and β_i , chooses a set X_i (and Y_i) and move on to the next item. After $dT^{2/3}\log(T)^{1/3}$ exploration rounds, RGL commits to the last chosen set X_d.

However, we argue that RGL explores too much, and that logarithmic, problem-dependent regret upper bounds can be obtained both in expectation and with high-probability.

3. Algorithm for full-bandit feedback: *Double-Greedy - Explore-then-Commit* (DG−ETC)

In this section, we propose *Double-Greedy - Explore-then-Commit* (DG−ETC), a novel algorithm for unconstrained submodular maximization (USM) with stochastic full bandit feedback. DG−ETC builds on insights from Buchbinder et al. (2012), Roughgarden and Wang (2018) and Harvey et al. (2020). We present the theoretical guarantees of DG−ETC in Section 4, which outperform existing algorithms for this setting.

In the following, the word *round* refers to a single increment of time t, the word *step* refers to the per-item exploration steps (containing several rounds) and the word *phase* refers to the exploration/exploitation phases (the exploration phase containing several per-item steps).

3.1. Algorithms presentation

DG−ETC is presented in Algorithm 2, and is built on two subroutines: DG−Sp (Algorithm 3) and UpdExp (Algorithm 4).

***Double-Greedy - Explore-then-Commit* (DG−ETC, Algorithm 2).** Algorithm DG−ETC is an algorithm implementing an *Explore-then-Commit* type strategy. It takes as inputs the set of items $\mathcal{D}$, $c > 0$ the range of f, the sub-Gaussian parameter of the noise $\sigma > 0$, as well as the horizon $T \in \mathbb{N}^*$ and a confidence level $\delta \in (0, 1)$.

It first performs d exploration steps (one per item in $\mathcal{D}$) [Lines 12 to 26], each lasting at most $4\tau_{\max}$ rounds where

$$\tau_{\max} = T^{2/3}\log(dT)^{1/3}\,. \tag{3}$$

Contrarily to RGL (Fourati et al., 2023), the duration of each exploration step is problem-adaptive, and can be considerably smaller than the worst case (See Section 5.3). It then spends the rest of the rounds [Lines 27 to 32] exploiting the collected information. During this phase, it does not play a fixed set, but repeatedly samples random sets based on d Bernoulli random variables with parameters $(p_j)_{j\in[d]}$ determined during the exploration phase.

***Double-Greedy - Sampling* (DG−Sp, Algorithm 3).** Both phases rely on the DG−Sp subroutine [Lines 15 and 30 in Algorithm 2], which is a variation of DG from Buchbinder et al. (2012) (Algorithm 1). DG−Sp relies on the parameters $(p_j)_{j\in[d]}$ provided by the meta-algorithm DG−ETC, which also provides an item $i \in [d+1]$ at which DG−Sp should stop. Like DG, it begins by initializing two sets X_0 and Y_0 as the empty and the full sets. Then it iterates over the parameters $(p_j)_{j\in[d]}$ and proceeds to either add (to X_j) or remove (from Y_j) in order to create $(X_j, Y_j)_{j<i}$ by sampling

Algorithm 2 Double-Greedy - Explore-then-Commit (DG-ETC)

Inputs: $\mathcal{D}$, $c > 0$, $\sigma > 0$, $\delta > 0$, $T \in \mathbb{N}^*$
/* Instantiating */
$d \leftarrow |\mathcal{D}|$
Instantiate $g_{T,\delta}$ and $\tau_{\max}$ with (4) and (3)
Instantiate UpdExp with $g_{T,\delta}$ and $\tau_{\max}$
/* Initialisation */
$(t,\ i) \leftarrow (1,\ 1)$
$(\hat{\alpha}_j,\ \hat{\beta}_j)_{j\in[d]} \leftarrow 0$
$(p_j)_{j\in[d]} \leftarrow 1/2$
$(\tau_j)_{j\in[d]} \leftarrow 0$
/* Exploration phase */
while $i \leq d$ **do**
  /* 4 rounds for item i */
  $(X_{i-1},\ Y_{i-1}) \leftarrow$ DG-Sp$\big(\mathcal{D},\ (p_j)_j,\ i\big)$
  Play:
    $A_t \leftarrow X_{i-1}$ $A_{t+1} \leftarrow X_{i-1} \cup \{i\}$
    $A_{t+2} \leftarrow Y_{i-1}$ $A_{t+3} \leftarrow Y_{i-1} \setminus \{i\}$
  Receive:
    $Z_t\ ,\ Z_{t+1}\ ,\ Z_{t+1}\ ,\ Z_{t+3}$
  Update:
    $\hat{\alpha}_i \leftarrow \big(\tau_i\hat{\alpha}_i + (Z_{t+1} - Z_t)\big)/(\tau_i + 1)$
    $\hat{\beta}_i \leftarrow \big(\tau_i\hat{\beta}_i + (Z_{t+3} - Z_{t+2})\big)/(\tau_i + 1)$
    $\tau_i \leftarrow \tau_i\ + 1$
    $(p_i,\ i) \leftarrow$ UpdExp$\big(i,\ (\hat{\alpha}_i,\ \hat{\beta}_i),\ \tau_i\big)$
    $t \leftarrow t + 4$
end while
/* Exploitation phase */
while $t \leq T$ **do**
  $(X_d,\ Y_d) \leftarrow$ DG-Sp$\big(\mathcal{D},\ (p_j)_j,\ i\big)$
  Play: $A_t \leftarrow X_d$
  Update: $t \leftarrow t + 1$
end while

Bernoulli random variables. At the end, DG-Sp returns $(X_{i-1},\ Y_{i-1})$ and DG-ETC then decides to either collect information when $i \leq d$ or exploit when $i = d + 1$. An example of sampling from DG-Sp is illustrated in Figure 1.

Exploration update for DG-ETC (UpdExp, Algorithm 4). During its exploration phase, DG-ETC calls the subroutine UpdExp [Line 24 in Algorithm 2]. The latter takes as inputs the index of the current step i, estimates of the marginal gains $(\alpha,\ \beta)$ and the current values of τ for item i. The objective of UpdExp is to check if we can determine an adequate Bernoulli parameter p for item i and/or if the exploration has lasted too long (if $\tau \geq \tau_{\max}$). In both those cases, UpdExp returns an adequate parameter p and index $i + 1$ to tell DG-ETC to switch to the next item. Otherwise, p stays the default $1/2$ and UpdExp returns current index i.

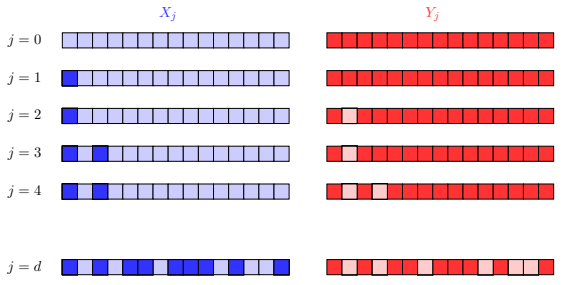


Figure 1: Example of sampling from DG-Sp, for $i = d + 1$ and $(K_j)_{j\in[d]} = (1, 0, 1, 0, \ldots 1)$.

3.2. Exploring just enough for zero exploitation regret: the key idea

In DG-ETC, the number of rounds devoted to the exploration for each item is adaptive, and is controlled by the subroutine UpdExp. Given estimated marginal gains $(\hat{\alpha}_i,\ \hat{\beta}_i)$ and an exploration time τ, UpdExp checks if it is possible to counterbalance the errors coming from the different sources of uncertainties, with high probability.

The per-round exploitation regret induced by all sources of uncertainty (estimations errors, random sampling, noise fluctuations) for item i, is bounded with high-probability (Proposition 2 in our analysis) by $g_{T,\delta}/\sqrt{\tau_i}$ where

$$g_{T,\delta} = \sqrt{2(2\sigma^2+c^2)}\sqrt{2\log(dT)+\log(1/\delta)}\left(1+2\sqrt{\tfrac{\log(dT)}{T}}+\tfrac{9c}{\sqrt{2\sigma^2+c^2}}\left(\tfrac{\log(dT)}{T}\right)^{1/3}\right). \quad (4)$$

On the other hand, the decision to either add or remove item i with probability p_i [Line 30 in DG-ETC (Alg. 2) and Lines 4-9 in DG-Sp (Alg. 3)] induces an average loss per exploration round of $l(\hat{\alpha}_i,\ \hat{\beta}_i,\ p_i)$ where

$$l(\alpha,\ \beta,\ p) = \max(l^+(\alpha,\ \beta,\ p), l^-(\alpha,\ \beta,\ p)), \quad (5)$$
$$\text{with } l^+(\alpha,\ \beta,\ p) = \big(1-p\big)\alpha - \tfrac{1}{2}(p\alpha+(1-p)\beta), \quad l^-(\alpha,\ \beta,\ p) = p\beta - \tfrac{1}{2}(p\alpha+(1-p)\beta).$$

In this definition, l^+ and l^- are per-round regrets of using parameter p when the estimated marginal gains are $(\alpha,\ \beta)$, corresponding to the two cases $\{i \in A^*\}$ and $\{i \notin A^*\}$. As one wants to hedge against both eventualities, we consider the worst-case loss l which is a max of both l^+ and l^-, which explains the form of Eq. (11).

Algorithm 3 *Double-Greedy - Sampling* (DG-Sp)

Inputs: $\mathcal{D}$, $(p_j)_{j\in[d]} \in [0,1]^d$, $i \in [d]$
$(X_0,\ Y_0) \leftarrow (\emptyset,\ \mathcal{D})$
for $j = 1, \dots, (i-1)$ **do**
  $K_j \sim \mathcal{B}(p_j)$
  if K_j **then**
    $(X_j,\ Y_j) \leftarrow (X_{j-1} \cup \{j\},\ Y_{j-1})$
  else
    $(X_j,\ Y_j) \leftarrow (X_{j-1},\ Y_{j-1} \setminus \{j\})$
  end if
end for
Return: $(X_{i-1},\ Y_{i-1})$

Algorithm 4 Exploration update (UpdExp)

Inputs: $i \in [d]$, $(\alpha,\ \beta) \in [-c,c]^2$, $\tau \in \mathbb{N}^*$
$\Lambda \leftarrow \{p \in [0,1] \text{ s. t. } l(\alpha,\ \beta,\ p) + g_{T,\delta}/\sqrt{\tau} \leq 0\}$
$p \leftarrow 1/2$
if $\Lambda = \emptyset$ **then**
  if $\tau \geq \tau_{\max}$ **then**
    $p \leftarrow \frac{\alpha_+}{\alpha_+ + \beta_+}$ where $(\cdot)_+ = \max\{\cdot, 0\}$
    $i \leftarrow i+1$
  end if
else
  $p \leftarrow p \in \Lambda$
  $i \leftarrow i+1$
end if
Return: $(p,\ i)$

UpdExp checks if, given estimations $(\hat{\alpha}_i,\ \hat{\beta}_i)$ and a current exploration duration $4\tau_i$, it is possible to find a parameter p_i so that the random errors $g_{T,\delta}/\sqrt{\tau_i}$ are absorbed by the loss $l(\hat{\alpha}_i,\ \hat{\beta}_i,\ p_i)$. Formally, it looks for the existence of a $p_i \in [0,1]$ so that

$$l(\hat{\alpha}_i,\ \hat{\beta}_i,\ p_i) + \tfrac{g_{T,\delta}}{\sqrt{\tau_i}} \leq 0\,, \quad (6)$$

which is guaranteed to happen after a logarithmic number of rounds (Proposition 3). If it is the case, UpdExp returns this parameter p_i and makes DG-ETC move on to the next item. Otherwise, the exploration for the current item i continues unless it has already lasted too long (i.e. if $\tau_i \geq \tau_{\max}$). In this case, UpdExp returns parameter $p_i = \frac{\alpha_+}{\alpha_+ + \beta_+}$ and makes DG-ETC move on to the next step. This last choice for p_i ensures the loss l to be negative (or null) in the exploitation phase and the per-round regret for item i to be bounded simply by $g_{T,\delta}/\sqrt{\tau_{\max}}$.

While RGL (Fourati et al., 2023) devotes the same number of rounds to all the items in the exploration phase, the subroutine UpdExp enables more flexibility. In particular, Section 4 links the number of exploration rounds necessary with problem-dependent quantities.

Remark 1 *The possibility to counterbalance the accumulated errors with negative losses is enabled by the approximate regret criterion, using the worst-case ratio, and an in-depth analysis of the original Double-Greedy algorithm. In all generality, this kind of intuition could also be applied to other methods to recover similar logarithmic upper bounds.*

4. Theoretical guarantees for DG-ETC

This section presents theoretical guarantees satisfied by our approach. We introduce a concept of problem-dependent *hardness* that characterizes how difficult it is to maximize a given submodular function with a *Double-Greedy* approach. We then show that DG-ETC satisfies logarithmic $1/2$-approximate pseudo-regret upper bounds which depend on this hardness, with a $O(dT^{2/3}\log(dT)^{1/3})$ worst-case.

We remind that the items are taken in an arbitrary order, and the quantities may depend on it.

4.1. Double-Greedy hardness

The following hardness notion relates to the sufficient number of exploration rounds that guarantee to find p_i's to induce zero $1/2$-approximate exploration regret.

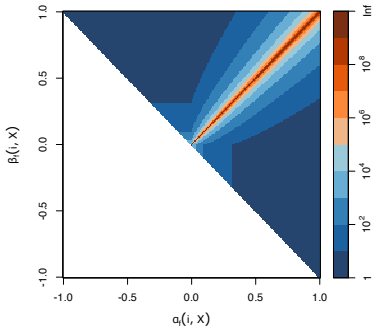
Definition 2 (DG-hardness) *Let $\mathcal{D}$ be a set of d elements, considered in an arbitrary order. Let f be a submodular set-function over $\mathcal{D}$ and i be an item in $\mathcal{D}$.*

We define the local DG-hardness *at item i as*

$$h_f(i) = \max_{X \subseteq [i-1]} \frac{\big(\alpha_f(i,X)_+ + \beta_f(i,X)_+\big)^2}{\big(\alpha_f(i,X)_+ - \beta_f(i,X)_+\big)^4},$$

where $(\,\cdot\,)_+ = \max\{\,\cdot\,, 0\}$ and

$$\alpha_f(i,X) = f(X \cup \{i\}) - f(X)\,,$$
$$\beta_f(i,X) = f\big((\mathcal{D} \setminus [i]) \cup X\big) - f\big((\mathcal{D} \setminus [i-1]) \cup X\big)\,.$$



We also define the global DG-hardness *as $H_f = \sum_{i \in [d]} h_f(i)$.*

Remark 2

- *This definition is actually not tight, as we will see in the analysis. We consider different configurations of $(\alpha,\ \beta)$, but we use this form as it is more convenient.*
- *We can also define a dual quantity, a* local DG-gap $\Delta_{f,i} = \big(h_f(i)\big)^{-1/2}$*, playing the same role as the suboptimality gaps in pseudo-regret upper bounds for stochastic multi-armed bandits (homogeneous to a difference of rewards). The corresponding* global DG-gap *is $\Delta_f = H_f^{-1/2}$.*

Example *For illustration purposes, let's consider the following function g: we assume there exists $\xi \in [-1,1]^d$ and $\nu \in (0,1]$ such that for all $X \subset [d]$,*

$$g(X) = \left(\sum_{i \in X,\, \xi_i \geq 0} \xi_i \right)^{\nu} - \left(\sum_{i \in X,\, \xi_i < 0} -\xi_i \right)^{1/\nu} + \|\xi_-\|_1^{1/\nu} \tag{7}$$

where $\xi_- = (\xi_i \mathbb{1}\{\xi_i < 0\})_i$ and the constant $\|\xi_-\|_1^{1/\nu}$ is there solely to guarantee the positivity of g. Then g is submodular and for all $i \in [d]$:

$$\Delta_{g,i} = \begin{cases} g([i]) - g([i-1]) & \text{if } \xi_i \geq 0\,, \\ g(\mathcal{D} \setminus [i]) - g(\mathcal{D} \setminus [i-1]) & \text{if } \xi_i < 0\,. \end{cases}$$

These expressions remind the notion of (local) suboptimality gaps common in bandit literature. If $\xi_i \geq 0$ then $i \in A^*$ and the DG-gap corresponds to the reward gained by adding i to $[i-1]$. If $\xi_i < 0$ then $i \notin A^*$ and the DG-gap corresponds to the reward increase when removing i from $\{i, i+1, \dots, d\}$. In particular, linear functions ($\nu = 1$) and functions defined as powers of the set cardinal ($\xi \in \{0,1\}^d$) can be written as in (7). If g is linear we have that $\Delta_{g,i} = \xi_i$ for all $i \in \mathcal{D}$.

4.2. Regret upper bounds for `DG-ETC`

We now present our main result, in the form of the following $1/2$-approximate pseudo-regret upper bounds for `DG-ETC`.

Theorem 2 *Let $\mathcal{D}$ be a set with $|\mathcal{D}| = d \in \mathbb{N}^*$, $T \in \mathbb{N}^*$ such that $d(T\sqrt{\log(dT)})^{2/3} \leq T/2$, $\sigma \in \mathbb{R}_+^*$ and $c \in \mathbb{R}_+^*$. Let $0 \in (0,1]$.*

Then, with probability greater than $1 - 10\delta/T$, `DG-ETC` satisfies

$$R_T \leq C_1 \min\left(H_f \log(dT), dT^{2/3} \log(dT)^{1/3} \right)$$

where C_1 is a constant independent from d, T and δ.

Likewise, in expectation,

$$\mathbb{E}[R_T] \leq C_2 \min\left(H_f \log(dT), dT^{2/3} \log(dT)^{1/3} \right),$$

where C_1 is a constant independent from d and T.

Remark 3 *We can get finer-grained bounds by using the local DG-hardnesses instead of the local one. From Eq.8 at the end of Section 5, we can keep the per-item granularity to get with probability at least $1 - 10\delta/T$*

$$R_T \leq C_3 \sum_{i \in [d]} \min\left(h_{f,i} \log(dT), T^{2/3} \log(dT)^{1/3} \right),$$

where C_1 is a constant independent from d, T and δ. In particular, depending on the scale of the horizon T with respect to the different local hardnesses $(h_{f,i})_{i \in [d]}$, we have sum of some logarithmic terms, and others of magnitude $T^{2/3} \log(dT)^{1/3}$.

5. Analysis of DG-ETC

This section presents a sketch of proof for Theorem 2.

We denote $\tau = \sum_{i\in[d]} 4\tau_i$ the last exploration round. For all the items i, we also denote $t_i = \sum_{j<i} 4\tau_i$, the last exploration round for item i. In this section, we have i-indices to denote items, and t-indices to denote that we place ourselves at the end of the round $t \in \mathbb{N}^*$. When t is not made explicit (notably for $\hat{\alpha}_i$, $\hat{\beta}_i$ and p_i), it means that we place ourselves after t_{i-1} and that they are fixed.

Outline of the proof. The idea of the proof is to find a high-probability event $\mathcal{E}$ under which the exploration phase takes a logarithmic number of rounds per-item, and the regret is non positive during the exploitation phase. To that end, we first breakdown the per-round regret of the exploitation phase into per-item contributions (Section 5.1). This enables to highlight $\mathcal{E}$ under which the per-round, per-item, regret is bounded by $l(\hat{\alpha}_i,\ \hat{\beta}_i,\ p_i) + g_{T,\delta}/\sqrt{\tau_i}$ for all the items i (Section 5.2). Lastly, we prove that under $\mathcal{E}$, for all the items and depending on the *DG-hardness* of f (Definition 2), a logarithmic number of exploration rounds is sufficient to find a weight p_i so that $l(\hat{\alpha}_i, \hat{\beta}_i, p_i) + g_{T,\delta}/\sqrt{\tau_i} \leq 0$. Additionally, at worst, `UpdExp` returns a p_i so that $l(\hat{\alpha}_i, \hat{\beta}_i, p_i) \leq 0$ when τ reaches $\tau_{\max}$ (Section 5.3).

Template bound. Let $\mathcal{E}$ be an event, defined later in Section 5.2. Then, the $1/2$-approximate pseudo-regret can be bounded as

$$R_T \leq \mathbb{1}\{\mathcal{E}^c\}\frac{cT}{2} + \mathbb{1}\{\mathcal{E}\}\left(2c\sum_{i=1}^{d}\tau_i + \frac{1}{2}\sum_{t=\tau+1}^{T} r_t\right). \tag{8}$$

where $r_t = f(A^*) - 2f(A_t)$. Under $\mathcal{E}^c$, we use the trivial upper bound $cT/2$ on the regret, and under $\mathcal{E}$, we separate the exploration and exploitation phases.

5.1. Double-Greedy breakdown: Per-item exploitation regrets

We use an approach similar to Buchbinder et al. (2012) to bound r_t with a sum of per-item terms.

Item-wise breakdown. Let $t > \tau$. We considers sets $(A^*_{i,t})_{i\in[d]}$, with $A^*_{0,t} = A^*$ and $A^*_{d,t} = A_t$, so that we can control the evolution of $(f(A_{i,t}))_{i\in[d]}$ from $f(A^*)$ to $f(A_t)$ using the coefficients $(\alpha_{i,t},\ \beta_{i,t})_{i\in[d]}$. We define

$$\begin{aligned}
&\text{For } i = 0, && A^*_{0,t} = A^*, && \text{with } X_{0,t} = \emptyset\,, && Y_{0,t} = \mathcal{D}\,,\\
&\forall i \in [d], && A^*_{i,t} = (A^* \cup X_{i,t}) \cap Y_{i,t}, && \text{with } X_{i,t} \subseteq A^*_{i,t} \subseteq Y_{i,t}\,, &&\\
&\text{For } i = d, && A^*_{d,t} = X_{d,t} = Y_{d,t} = A_t\,, &&
\end{aligned} \tag{9}$$

where $X_{i,t} = \{j \leq i,\ K_{j,t} = 1\}$ and $Y_{i,t} = \mathcal{D} \setminus \{j \leq i,\ K_{j,t} = 0\}$ are the sets defined in `DG-Sp`.

Using these sets and the definition of r_t in Eq. (8), a telescopic argument yields

$$\begin{aligned}
r_t &\leq f(A^*) - f(A_t) - \frac{1}{2}\Big[2f(A_t) - (f(\emptyset) + f(\mathcal{D}))\Big] && \leftarrow (f \geq 0)\\
&= \Big[f(A^*_{0,t}) - f(A^*_{d,t})\Big] - \frac{1}{2}\Big[f(X_{d,t}) - f(X_{0,t}) + f(Y_{d,t}) - f(Y_{0,t})\Big]\\
&= \sum_{i=1}^{d}\Big[f(A^*_{i-1,t}) - f(A^*_{i,t}) - \frac{1}{2}\big(K_{i,t}\alpha_{i,t} + (1 - K_{i,t})\beta_{i,t}\big)\Big]\,,
\end{aligned} \tag{10}$$

where for all $i \in [d]$, $\alpha_{i,t} = f(X_{i-1,t} \cup \{i\}) - f(X_{i-1,t})$ and $\beta_{i,t} = f(Y_{i-1,t} \setminus \{i\}) - f(Y_{i-1,t})$.

Submodularity. While the marginal gains $(\alpha_{i,t},\ \beta_{i,t})_{i\in[d]}$ could be estimated, the sets A^*, and $(A^*_{i,t})_{i\in[d]}$ remain unknown. However, the definition of $(A^*_{i,t})_{i\in[d]}$ and submodularity yield

- If $[i \in A^*]$, then $f(A^*_{i-1,t}) - f(A^*_{i,t}) \leq (1 - K_{i,t})\alpha_{i,t}$;
- Else $[i \notin A^*]$, and $f(A^*_{i-1,t}) - f(A^*_{i,t}) \leq K_{i,t}\beta_{i,t}$.

Eq. (10) then becomes

$$r_t \leq \sum_{i\in[d]} \left[\mathbb{1}\{i \in A^*\}(1 - K_{i,t})\alpha_{i,t} + \mathbb{1}\{i \notin A^*\}K_{i,t}\beta_{i,t} - \frac{1}{2}\big(K_{i,t}\alpha_{i,t} + (1 - K_{i,t})\beta_{i,t}\big)\right].$$

Since $[i \in A^*]$ and $[i \notin A^*]$ are exclusive, we have

$$\sum_{t=\tau+1}^{T} r_t \leq \sum_{i\in[d]} \max\left(R^+_{T,i},\ R^-_{T,i}\right), \tag{11}$$

where

$$R^+_{T,i} = \textstyle\sum_{t=\tau+1}^{T} \left[(1 - K_{i,t})\alpha_{i,t} - \frac{1}{2}\big(K_{i,t}\alpha_{i,t} + (1 - K_{i,t})\beta_{i,t}\big)\right],$$
$$R^-_{T,i} = \textstyle\sum_{t=\tau+1}^{T} \left[K_{i,t}\beta_{i,t} - \frac{1}{2}\big(K_{i,t}\alpha_{i,t} + (1 - K_{i,t})\beta_{i,t}\big)\right].$$

5.2. High-probability exploitation regret

Let $i \in [d]$, the objective now is to control $\max\left(R^+_{T,i},\ R^-_{T,i}\right)$ from Eq. (11). To that end, the following propositions (proven in Appendix C.1) states how the errors coming from the different randomness sources concentrate.

Proposition 2 *Let $\mathcal{H}$ and $\mathcal{E}$ be the event*

$$\mathcal{H} = \left\{\begin{array}{ll} \forall i \in [d],\ \forall t > t_{i-1}, & |\bar{\alpha}_i - \hat{\alpha}_{i,t}| \leq \sqrt{2\sigma^2 + c^2}\sqrt{2\frac{\log(dT/\delta)+\log(1+4\tau_{i,t})}{\tau_{i,t}+1}}\ ; \\ & |\bar{\beta}_i - \hat{\beta}_{i,t}| \leq \sqrt{2\sigma^2 + c^2}\sqrt{2\frac{\log(dT/\delta)+\log(1+4\tau_{i,t})}{\tau_{i,t}+1}} \end{array}\right\},$$

$$\mathcal{E} = \mathcal{H} \cap \left\{\forall i \in [d], \quad \max\left(R^+_{T,i}, R^-_{T,i}\right) - (T - \tau)\Big(l(\hat{\alpha}_i,\ \hat{\beta}_i,\ p_i) + g_{T,\delta}/\sqrt{\tau_i}\Big) \leq 0\right\},$$

where for all $i \in [d]$, $\bar{\alpha}_i = \mathbb{E}\big[\alpha_{i,t}|(p_j)_{j<i}\big]$ and $\bar{\beta}_i = \mathbb{E}\big[\beta_{i,t}|(p_j)_{j<i}\big]$, both quantities being constant for rounds $t > t_{i-1}$, and $g_{T,\delta}$ is defined in Eq. (4).

Then, $\mathbb{P}(\mathcal{H}^c) \leq \frac{4\delta}{T}$, and $\mathbb{P}(\mathcal{E}^c) \leq \frac{10\delta}{T}$.

Template bound. Reinjecting Eq. (11) and Proposition 2 yields

$$R_T \leq \mathbb{1}\{\mathcal{E}^c\}\frac{cT}{2} + \mathbb{1}\{\mathcal{E}\}\sum_{i\in[d]}\left(2c\tau_i + (T - \tau)\Big(l(\hat{\alpha}_i, \hat{\beta}_i, p_i) + \frac{g_{T,\delta}}{\sqrt{\tau_i}}\Big)\right), \tag{12}$$

where $\mathcal{E}$ is the event defined Proposition 2.

5.3. Sufficient exploration

In this section, we analyze the exploration steps for each item. We exhibit sufficient conditions for the exploration to only last a logarithmic number of rounds, the choice of p_i when $\tau_i \geq \tau_{\max}$ ensuring a $(O(T\sqrt{\log(dT)})^{2/3})$ regret for item i as the average loss would be non positive. Subroutine `UpdExp` looks for a parameter $p \in [0,1]$ so that

$$\big(1-p\big)\hat{\alpha}_i - \frac{1}{2}\big(p\hat{\alpha}_i + (1-p)\hat{\beta}_i\big) \leq -\frac{g_{T,\delta}}{\sqrt{\tau_i}}, \quad \text{and} \quad p\hat{\beta}_i - \frac{1}{2}\big(p\hat{\alpha}_i + (1-p)\hat{\beta}_i\big) \leq -\frac{g_{T,\delta}}{\sqrt{\tau_i}}\,. \tag{13}$$

Under $\mathcal{E}$, as we can upper bound $|\hat{\alpha}_i - \bar{\alpha}_i|$ and $|\hat{\beta}_i - \bar{\beta}_i|$, it is sufficient to have

$$\begin{cases} \Big(1-p\Big)\bar{\alpha}_i - \frac{1}{2}\big(p\bar{\alpha}_i + (1-p)\bar{\beta}_i\big) & \leq -\frac{g_{T,\delta}}{\sqrt{\tau_i}} - \frac{3}{2}\sqrt{2\sigma^2+c^2}\sqrt{2\frac{\log(dT/\delta)+\log(1+4\tau_i)}{\tau_i+1}} \\ p\bar{\beta}_i - \frac{1}{2}\big(p\bar{\alpha}_i + (1-p)\bar{\beta}_i\big) & \leq -\frac{g_{T,\delta}}{\sqrt{\tau_i}} - \frac{3}{2}\sqrt{2\sigma^2+c^2}\sqrt{2\frac{\log(dT/\delta)+\log(1+4\tau_i)}{\tau_i+1}}\,, \end{cases}$$

which in turn is implied by

$$p(\bar{\beta}_i - 3\bar{\alpha}_i) \leq -\frac{g_i + \gamma_{T,\delta}}{\sqrt{\tau_i}} + (\beta_i - 2\bar{\alpha}_i), \qquad p(3\bar{\beta}_i - \bar{\alpha}_i) \leq -\frac{g_i + \gamma_{T,\delta}}{\sqrt{\tau_i}} + \bar{\beta}_i\,, \tag{14}$$

where $\gamma_{T,\delta} = 3\sqrt{(2\sigma^2+c^2)(\log(dT/\delta) + \log(1+T))}$.

The following proposition gives sufficient conditions to find a p_i for Eq. (13) to be satisfied.

Proposition 3 *For each items i, under event $\mathcal{E}$ defined in Proposition 2, `UpdExp` finds a weight p_i such that $l(\hat{\alpha}_{i,t}, \hat{\beta}_{i,t}, p_i) + g_{T,\delta}/\sqrt{\tau_{i,t}} \leq 0$ before $\tau_{i,t}$ has reached $(g_{T,\delta} + \gamma_{T,\delta})^2\, h_{f,i}$.*

Template bound. Using Proposition 3, the upper bound Eq. (12) becomes

$$\begin{aligned} R_T &\leq \mathbb{1}\{\mathcal{E}^c\}\frac{cT}{2} + \mathbb{1}\{\mathcal{E}\}\sum_{i\in[d]}\Bigg[2c\min\Big((g_{T,\delta}+\gamma_{T,\delta})^2 h_f(i),\ \tau_{\max}\Big) \\ &\qquad\qquad + \mathbb{1}\Big\{(g_{T,\delta}+\gamma_{T,\delta})^2 h_f(i) > \tau_{\max}\Big\}\tau_{\max}\frac{T g_{T,\delta}}{(\tau_{\max})^{3/2}}\Bigg] \\ &= \mathbb{1}\{\mathcal{E}^c\}\frac{cT}{2} + \mathbb{1}\{\mathcal{E}\}\sum_{i\in[d]}\left(2c + \frac{g_{T,\delta}}{\log(dT)^{1/2}}\right)\min\Big((g_{T,\delta}+\gamma_{T,\delta})^2 h_f(i),\ \tau_{\max}\Big)\,. \end{aligned} \tag{15}$$

Event $\mathcal{E}$ happens with probability greater than $1 - \frac{10\delta}{T}$ (Proposition 2), thus the high-probability result. Choosing $\delta = 1$ yields the bound in expectation.

6. Concluding remarks

We propose and analyze Algorithm `DG-ETC` for the online unconstrained submodular maximization problem with stochastic bandit feedback. Our algorithm is a considerable improvement from other existing approaches, as it satisfies logarithmic upper bounds for the $1/2$-approximate pseudo-regret, dependant on a new notion of hardness that we introduce. Possible extensions include designing anytime variants, and algorithms adaptive to the adversarial/stochastic setting (best of both worlds).

An interesting feature of `DG-ETC` is that is leverages the looseness of worst-case approximation ratios in non-adversarial cases, and we argue that this kind of strategy could also be applied to other settings to yields similar performances.

References

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. *Advances in Neural Information Processing Systems*, 24, 2011.

Francis Bach. Learning with submodular functions: A convex optimization perspective. *Foundations and Trends® in Machine Learning*, 6(2-3):145–373, 2013.

Sébastien Bubeck, Nicolo Cesa-Bianchi, and Sham M Kakade. Towards minimax policies for online linear optimization with bandit feedback. In *Conference on Learning Theory*, pages 41–1. JMLR Workshop and Conference Proceedings, 2012.

Niv Buchbinder, Moran Feldman, Joseph (Seffi) Naor, and Roy Schwartz. A tight linear time (1/2)-approximation for unconstrained submodular maximization. In *Proceedings of the 2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*. IEEE Computer Society, 2012.

Wei Chen, Yajun Wang, and Yang Yuan. Combinatorial multi-armed bandit: General framework and applications. In *International Conference on Machine Learning*, 2013.

Rémy Degenne and Vianney Perchet. Combinatorial semi-bandit with known covariance. *Advances in Neural Information Processing Systems*, 2016.

Louis Faury, Marc Abeille, Clément Calauzènes, and Olivier Fercoq. Improved optimistic algorithms for logistic bandits. In *International Conference on Machine Learning*, 2020.

Uriel Feige, Vahab S Mirrokni, and Jan Vondrák. Maximizing non-monotone submodular functions. *SIAM Journal on Computing*, 40(4):1133–1153, 2011.

Fares Fourati, Vaneet Aggarwal, Christopher John Quinn, and Mohamed-Slim Alouini. Randomized greedy learning for non-monotone stochastic submodular maximization under full-bandit feedback. In *International Conference on Artificial Intelligence and Statistics*, 2023.

Yanjun Han, Yining Wang, and Xi Chen. Adversarial combinatorial bandits with general non-linear reward functions. In *International Conference on Machine Learning*, 2021.

Nicholas Harvey, Christopher Liaw, and Tasuku Soma. Improved algorithms for online submodular maximization via first-order regret bounds. *Advances in Neural Information Processing Systems*, 2020.

Shinji Ito. Hybrid regret bounds for combinatorial semi-bandits and adversarial linear bandits. *Advances in Neural Information Processing Systems*, 2021.

Branislav Kveton, Zheng Wen, Azin Ashkan, and Csaba Szepesvári. Tight regret bounds for stochastic combinatorial semi-bandits. In *International Conference on Artificial Intelligence and Statistics*, 2015.

Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

L. Lovász. *Submodular functions and convexity*, pages 235–257. Springer Berlin Heidelberg, Berlin, Heidelberg, 1983.

George L Nemhauser, Laurence A Wolsey, and Marshall L Fisher. An analysis of approximations for maximizing submodular set functions–I. *Mathematical Programming*, 14:265–294, 1978.

Gergely Neu and Michal Valko. Online combinatorial optimization with stochastic decision sets and adversarial losses. *Advances in Neural Information Processing Systems*, 2014.

Rad Niazadeh, Negin Golrezaei, Joshua R Wang, Fransisca Susan, and Ashwinkumar Badanidiyuru. Online learning via offline greedy algorithms: Applications in market design and optimization. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, 2021.

Pierre Perrault, Etienne Boursier, Michal Valko, and Vianney Perchet. Statistical efficiency of thompson sampling for combinatorial semi-bandits. *Advances in Neural Information Processing Systems*, 2020.

Tim Roughgarden and Joshua R. Wang. An optimal algorithm for online unconstrained submodular maximization. In *Conference On Learning Theory*, 2018.

Martin J Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press, 2019.

Julien Zhou, Pierre Gaillard, Thibaud Rahier, Houssam Zenati, and Julyan Arbel. Towards efficient and optimal covariance-adaptive algorithms for combinatorial semi-bandits. *Advances in Neural Information Processing Systems*, 2024.

Appendix A. Sub-Gaussianity

We use sub-Gaussianity assumptions and use common concentration tools to control the noise from $(\eta_t)_{t\in[T]}$ and the randomization from the algorithm we propose. This section remind the results that we use.

Definition 3 (Sub-Gaussian) *Let $\sigma > 0$ and X be a real-valued random variable such that $\mathbb{E}[X] = 0$. We say that X is σ^2-sub-Gaussian, for all $\lambda \in \mathbb{R}$,*

$$\mathbb{E}[\exp(\lambda X)] \leq \exp\left(\frac{\lambda^2\sigma^2}{2}\right).$$

In particular, for bounded independent random variables, we have the following lemma.

Lemma 1 (Hoeffding's inequality for sum of i.i.d. bounded r.v.) *Let $\delta > 0$, $N \in \mathbb{N}^*$, and $(Z_n)_{n\in[N]}$ a family of i.i.d. real random variables bounded in $[a, b]$ where $(a, b) \in (\mathbb{R})^2$, with mean $\mu \in [a, b]$.*

Then for all $n \in [N]$, Z_n is $\frac{(b-a)^2}{4}$-sub-Gaussian, and with probability at least $1 - \delta$,

$$\frac{1}{N}\sum_{n=1}^{N}\left[Z_n - \mu\right] < \frac{b-a}{2}\sqrt{\frac{2}{N}\log(1/\delta)}\,.$$

The sub-Gaussianity for bounded random variables an the concentration for the sums of i.i.d random variables are classical results proven that can be found Wainwright (2019) for example.

As we estimate quantities in an online setting, with observations arriving sequentially and depending on our actions, we need a more powerful tool. This is provided by the following lemma.

Lemma 2 (Hoeffding's inequality with martingales) *Let $\delta > 0$, $\sigma > 0$. Let $(\mathcal{G}_t)_{t\in\mathbb{N}}$ be a filtration and $(Z_t)_{t\in\mathbb{N}^*}$ a $(\mathcal{G}_t)$-adapted martingale with $\mathbb{E}[Z_1] = 0$. We assume that for all $t \in \mathbb{N}$, Z_{t+1} is σ^2-sub-Gaussian conditionally to $\mathcal{G}_t$. Let $(U_t)_{t\in\mathbb{N}^*}$ be a $(\mathcal{G}_t)$-predictable process. Then, with probability at least $1 - \delta$, for all $t \in \mathbb{N}$*

$$\frac{\sum_{s=1}^{t} U_s Z_s}{1 + \sum_{s=1}^{t} U_s^2} < \frac{\sigma}{\sqrt{1 + \sum_{s=1}^{t} U_s^2}}\sqrt{2\log(1/\delta) + \log\left(1 + \sum_{s=1}^{t} U_s^2\right)}$$

The proof relies on the method of mixture, widely used in the bandit literature (Abbasi-Yadkori et al., 2011; Faury et al., 2020; Zhou et al., 2024).

Proof Let $\delta > 0, \sigma > 0$. Let $(\mathcal{G}_t)$ be a filtration and (Z_t) be a $\mathcal{G}_t$-adapted martingale with $\mathbb{E}[Z_1] = 0$ and so that for all $t \in \mathbb{N}$, Z_{t+1} is σ^2-sub-Gaussian conditionally to $\mathcal{G}_t$. Let (U_t) be a $\mathcal{G}_t$-predictable process.

Let $t \in \mathbb{N}^*$, a first direct result is that, $U_t Z_t$ is $(\sigma U_t)^2$-sub-Gaussian conditionally to $\mathcal{G}_{t-1}$. Let $\lambda \in \mathbb{R}$. Then,

$$\mathbb{E}\Big[\exp\left(\lambda U_t Z_t - \frac{\lambda^2}{2}(\sigma U_t)^2\right)|\mathcal{G}_{t-1}\Big] \leq 1\,. \tag{16}$$

We define

$$M_t(\lambda) = \exp\left(\lambda \sum_{s=1}^{t} U_s Z_s - \frac{\lambda^2}{2} \sum_{s=1}^{t} (\sigma U_s)^2\right)$$

with $M_0(\lambda) = 1$. From eq. (16),

$$\begin{aligned}
\forall t \in \mathbb{N}, \qquad \mathbb{E}[M_t(\lambda)|\mathcal{G}_t] &= \mathbb{E}\left[\exp\left(\lambda \sum_{s=1}^{t} U_s Z_s - \frac{\lambda^2}{2} \sum_{s=1}^{t} (\sigma U_s)^2\right)\middle| \mathcal{G}_{t-1}\right] \\
&= M_{t-1}(\lambda)\, \mathbb{E}\left[\exp\left(\lambda U_t Z_t - \frac{\lambda^2}{2}(\sigma U_t)^2\right)\middle| \mathcal{G}_{t-1}\right] \\
&\leq M_{t-1}(\lambda)\,.
\end{aligned}$$

Then, $(M_t(\lambda))_t$ is a $\mathcal{G}_t$-supermartingale, with $\mathbb{E}[M_t(\lambda)] \leq 1$.

We now consider $\lambda \sim \mathcal{N}(0, 1/\sigma^2)$, independent from all the other distributions, then we can define

$$\begin{aligned}
\bar{M}_t &= \mathbb{E}_{\lambda\sim\mathcal{N}(0,1/\sigma^2)}[M_t(\lambda)] \\
&= \frac{\sigma}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\Big(-\frac{(\sigma x)^2}{2}\Big) \exp\Big(x \sum_{s=1}^{t} U_s Z_s - \frac{x^2}{2} \sum_{s=1}^{t} (\sigma U_s)^2\Big) dx \\
&= \frac{\sigma}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\Big(-\frac{(\sigma x)^2(1+\sum_{s=1}^{t} U_s^2)}{2} + x \sum_{s=1}^{t} U_s Z_s\Big) dx \\
&= \frac{\sigma}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\left(-\frac{\sigma^2(1+\sum_{s=1}^{t} U_s^2)}{2}\Big(x^2 - 2x \frac{\sum_{s=1}^{t} U_s Z_s}{\sigma^2(1+\sum_{s=1}^{t} U_s^2)}\Big)\right) dx \\
&= \frac{\sigma}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\left(-\frac{\sigma^2(1+\sum_{s=1}^{t} U_s^2)}{2}\Big(x - \frac{\sum_{s=1}^{t} U_s Z_s}{\sigma^2(1+\sum_{s=1}^{t} U_s^2)}\Big)^2 + \frac{(\sum_{s=1}^{t} U_s Z_s)^2}{2\sigma^2(1+\sum_{s=1}^{t} U_s^2)}\right) dx \\
&= \exp\left(\frac{(\sum_{s=1}^{t} U_s Z_s)^2}{2\sigma^2(1+\sum_{s=1}^{t} U_s^2)}\right) \frac{\sigma}{\sqrt{2\pi}} \frac{\sqrt{2\pi}}{\sigma\sqrt{1+\sum_{s=1}^{t} U_s^2}} \frac{\sigma\sqrt{1+\sum_{s=1}^{t} U_s^2}}{\sqrt{2\pi}} \\
&\qquad\qquad \int_{\mathbb{R}} \exp\left(-\frac{\sigma^2(1+\sum_{s=1}^{t} U_s^2)}{2}\Big(x - \frac{\sum_{s=1}^{t} U_s Z_s}{\sigma^2(1+\sum_{s=1}^{t} U_s^2)}\Big)^2\right) dx \\
&= \exp\left(\frac{(\sum_{s=1}^{t} U_s Z_s)^2}{2\sigma^2(1+\sum_{s=1}^{t} U_s^2)}\right) \frac{1}{\sqrt{1+\sum_{s=1}^{t} U_s^2}} \mathbb{E}_{\lambda\sim\mathcal{N}\left(\frac{\sum_{s=1}^{t} U_s Z_s}{\sigma^2(1+\sum_{s=1}^{t} U_s^2)},\, \frac{1}{\sigma^2(1+\sum_{s=1}^{t} U_t^2)}\right)}\Big[1\Big] \\
&= \frac{1}{\sqrt{1+\sum_{s=1}^{t} U_s^2}} \exp\left(\frac{(\sum_{s=1}^{t} U_s Z_s)^2}{2\sigma^2(1+\sum_{s=1}^{t} U_s^2)}\right) \\
\bar{M}_t &= \exp\left(\frac{(\sum_{s=1}^{t} U_s Z_s)^2}{2\sigma^2(1+\sum_{s=1}^{t} U_s^2)} - \frac{1}{2}\log\Big(1+\sum_{s=1}^{t} U_s^2\Big)\right)
\end{aligned}$$

Besides,

$$\begin{aligned}\mathbb{E}\Big[\bar{M}_t\Big|\mathcal{G}_{t-1}\Big] &= \mathbb{E}\Big[\mathbb{E}_{\lambda\sim\mathcal{N}(0,1/\sigma^2)}[M_t(\lambda)]\Big|\mathcal{G}_{t-1}\Big] \\ &= \mathbb{E}_{\lambda\sim\mathcal{N}(0,1/\sigma^2)}\Big[\mathbb{E}[M_t(\lambda)|\mathcal{G}_{t-1}]\Big] \\ &\leq \mathbb{E}_{\lambda\sim\mathcal{N}(0,1/\sigma^2)}\Big[M_{t-1}(\lambda)\Big] \\ &= \bar{M}_{t-1}\,.\end{aligned}$$

So $(\bar{M}_t)_t$ is also a supermartingale, which yield that

$$\mathbb{E}[\bar{M}_t] \leq \mathbb{E}[\bar{M}_0] = 1\,.$$

Let $u_t > 0$. Now, using Chernoff's method,

$$\begin{aligned}\mathbb{P}\left(\frac{\sum_{s=1}^t U_s Z_s}{1+\sum_{s=1}^t U_s^2} \geq u_t\right) &\leq \mathbb{P}\left(\exp\Big(\frac{(\sum_{s=1}^t U_s Z_s)^2}{2\sigma^2(1+\sum_{s=1}^t U_s^2)} - \frac{u_t^2}{2\sigma^2}(1+\sum_{s=1}^t U_s^2)\Big) \geq 1\right) \\ &\leq \mathbb{E}\left[\exp\Big(\frac{(\sum_{s=1}^t U_s Z_s)^2}{2\sigma^2(1+\sum_{s=1}^t U_s^2)} - \frac{u_t^2}{2\sigma^2}(1+\sum_{s=1}^t U_s^2)\Big)\right] \\ &\leq \mathbb{E}\left[\bar{M}_t \exp\Big(\frac{1}{2}\log(1+\sum_{s=1}^t U_s^2) - \frac{u_t^2}{2\sigma^2}(1+\sum_{s=1}^t U_s^2)\Big)\right].\end{aligned}$$

Choosing $u_t = \frac{\sigma}{\sqrt{1+\sum_{s=1}^t U_s^2}}\sqrt{2\log(1/\delta) + \log(1+\sum_{s=1}^t U_s^2)}$,

$$\begin{aligned}\mathbb{P}\left(\frac{\sum_{s=1}^t U_s Z_s}{1+\sum_{s=1}^t U_s^2} \geq u_t\right) &\leq \mathbb{E}\Big[\delta\bar{M}_t\Big] \\ &\leq \delta\,.\end{aligned}$$

The bound for all t is based on the stopping time construction from Abbasi-Yadkori et al. (2011).

■

Appendix B. Proof for the high-probability bound of DG

Proposition 1 *Let $\mathcal{D}$ be a finite set, $\delta > 0$ and $T \in \mathbb{N}^*$ such that $T > 2\log(1/\delta)$. If $(S_i)_{i\in[T]}$ is the sequence of sets obtained by running independently T times DG Algorithm, then*

$$\max_{i\in[T]} f(S_i) > \Big(\frac{1}{2} - \frac{\log(1/\delta)}{T}\Big) f(A^*)\,, \qquad \textit{w.p.} \quad 1-\delta\,.$$

Proof Let $1 > \delta > 0$ and $T \in \mathbb{N}^*$, $T > 2\log(1/\delta)$. Then $(f(S_i))_i \in [T]$ is a sequence of T i.i.d. random variables, bounded in $[0, f(A^*)]$.

Let $\frac{1}{2} > u > 0$. Then

$$\begin{aligned}
\mathbb{P}\left(\max_{i\in[T]} f(S_i) < \Big(\frac{1}{2}-u\Big)f(A^*)\right) &= \mathbb{P}\left(\forall i \in [T],\ f(S_i) < \Big(\frac{1}{2}-u\Big)f(A^*)\right) \\
&= \prod_{i=1}^{T} \mathbb{P}\left(f(S_i) < \Big(\frac{1}{2}-u\Big)f(A^*)\right) \\
&\leq \mathbb{P}\left(\Big(\frac{1}{2}-u\Big)f(A^*) + f(A^*) - f(S_1) > f(A^*)\right)^T \\
&\leq \frac{1}{f(A^*)^T}\mathbb{E}\left[\Big(\frac{1}{2}-u\Big)f(A^*) + f(A^*) - f(S_1)\right]^T \leftarrow \text{Markov} \\
&\leq \frac{1}{f(A^*)^T}\left[\Big(\frac{1}{2}-u\Big)f(A^*) + \frac{1}{2}f(A^*)\right]^T \leftarrow \text{Theorem 1} \\
&= (1-u)^T \\
&\leq \exp(-Tu)
\end{aligned}$$

Therefore, taking $u = \frac{\log(1/\delta)}{T}$, we have the result

$$\mathbb{P}\left(\max_{i\in[T]} f(S_i) < \Big(\frac{1}{2}-u\Big)f(A^*)\right) \leq \delta\,.$$

■

Appendix C. Proofs for the analysis of DG-ETC

C.1. Proof for the high-probability exploitation regret

Proposition 2 *Let $\mathcal{H}$ and $\mathcal{E}$ be the event*

$$\mathcal{H} = \left\{\forall i \in [d],\ \forall t > t_{i-1},\quad \begin{array}{l} |\bar{\alpha}_i - \hat{\alpha}_{i,t}| \leq \sqrt{2\sigma^2 + c^2}\sqrt{2\frac{\log(dT/\delta)+\log(1+4\tau_{i,t})}{\tau_{i,t}+1}}\,; \\ |\bar{\beta}_i - \hat{\beta}_{i,t}| \leq \sqrt{2\sigma^2 + c^2}\sqrt{2\frac{\log(dT/\delta)+\log(1+4\tau_{i,t})}{\tau_{i,t}+1}} \end{array}\right\},$$

$$\mathcal{E} = \mathcal{H} \cap \left\{\forall i \in [d],\quad \max\left(R^+_{T,i}, R^-_{T,i}\right) - (T-\tau)\Big(l(\hat{\alpha}_i,\ \hat{\beta}_i,\ p_i) + g_{T,\delta}/\sqrt{\tau_i}\Big) \leq 0\right\},$$

where for all $i \in [d]$, $\bar{\alpha}_i = \mathbb{E}\big[\alpha_{i,t}|(p_j)_{j<i}\big]$ and $\bar{\beta}_i = \mathbb{E}\big[\beta_{i,t}|(p_j)_{j<i}\big]$, both quantities being constant for rounds $t > t_{i-1}$, and $g_{T,\delta}$ is defined in Eq. (4).

Then, $\mathbb{P}(\mathcal{H}^c) \leq \frac{4\delta}{T}$, and $\mathbb{P}(\mathcal{E}^c) \leq \frac{10\delta}{T}$.

Proof

We remind Eq. (11),

$$\sum_{t=\tau+1}^{T} r_t \leq \sum_{i\in[d]} \max\left(R^+_{T,i},\ R^-_{T,i}\right), \tag{11}$$

where

$R^+_{T,i} = \sum_{t=\tau+1}^{T}\Big[(1-K_{i,t})\alpha_{i,t} - \frac{1}{2}\big(K_{i,t}\alpha_{i,t} + (1-K_{i,t})\beta_{i,t}\big)\Big]$,

$R^-_{T,i} = \sum_{t=\tau+1}^{T}\Big[K_{i,t}\beta_{i,t} - \frac{1}{2}\big(K_{i,t}\alpha_{i,t} + (1-K_{i,t})\beta_{i,t}\big)\Big]$.

For all $i \in [d]$, we define $\bar{\alpha}_i = \mathbb{E}\big[\alpha_{i,t}|(p_j)_{j<i}\big]$ and $\bar{\beta}_i = \mathbb{E}\big[\beta_{i,t}|(p_j)_{j<i}\big]$, both quantities being constant for rounds $t > t_{i-1} \geq \tau$,

Separating the different sources of randomness yields

$$R^+_{T,i} = \bar{E}^+_{T,i} + \hat{E}^+_{T,i} + L^+_{T,i}\,, \qquad R^-_{T,i} = \bar{E}^-_{T,i} + \hat{E}^-_{T,i} + L^-_{T,i}\,.$$

where we have

- errors coming from the deviation of $(\alpha_{i,t},\ \beta_{i,t})$ from $(\bar{\alpha}_i,\ \bar{\beta}_i)$

  $\bar{E}^+_{T,i} = \sum_{t=\tau+1}^{T}\Big[(1-K_{i,t})(\alpha_{i,t}-\bar{\alpha}_i) - \frac{1}{2}\big(K_{i,t}(\alpha_{i,t}-\bar{\alpha}_i) + K^c_{i,t}(\beta_{i,t}-\bar{\beta}_i)\big)\Big]$,

  $\bar{E}^-_{T,i} = \sum_{t=\tau+1}^{T}\Big[K_{i,t}(\beta_{i,t}-\bar{\beta}_i) - \frac{1}{2}\big(K_{i,t}(\alpha_{i,t}-\bar{\alpha}_i) + K^c_{i,t}(\beta_{i,t}-\bar{\beta}_i)\big)\Big]$,

- approximation errors for $(\hat{\alpha}_i, \hat{\beta}_i)$:

  $\hat{E}^+_{T,i} = \sum_{t=\tau+1}^{T}\Big[(1-K_{i,t})(\bar{\alpha}_{i,t}-\hat{\alpha}_i) - \frac{1}{2}\big(K_{i,t}(\bar{\alpha}_{i,t}-\hat{\alpha}_i) + (1-K_{i,t})(\bar{\beta}_{i,t}-\hat{\beta}_i)\big)\Big]$,

  $\hat{E}^-_{T,i} = \sum_{t=\tau+1}^{T}\Big[K_{i,t}(\bar{\beta}_{i,t}-\hat{\beta}_i) - \frac{1}{2}\big(K_{i,t}(\bar{\alpha}_{i,t}-\hat{\alpha}_i) + (1-K_{i,t})(\bar{\beta}_{i,t}-\hat{\beta}_i)\big)\Big]$,

- the deviation of losses caused by the randomization of $(K_{i,t})_{i,t}$'s:

  $L^+_{T,i} = \sum_{t=\tau+1}^{T}(1-K_{i,t})\hat{\alpha}_i - \frac{1}{2}(K_{i,t}\hat{\alpha}_i + (1-K_{i,t})\hat{\beta}_i) - (T-\tau)l^+_i$,

  $L^-_{T,i} = \sum_{t=\tau+1}^{T}(K_i)\hat{\beta}_i - \frac{1}{2}(K_i, t\hat{\alpha}_i + (1-K_{i,t})\hat{\beta}_i) - (T-\tau)l^-_i$,

- the average loss criterion used in `UpdExp`:

  $l^+_i = l^+(\hat{\alpha}_i,\ \hat{\beta}_i,\ p_i) = (1-p_i)\hat{\alpha}_i - \frac{1}{2}(p_i\hat{\alpha}_i + (1-p_i)\hat{\alpha}_i)$,

  $l^-_i = l^-(\hat{\alpha}_i,\ \hat{\beta}_i,\ p_i) = p_i\hat{\beta}_i - \frac{1}{2}(p_i\hat{\alpha}_i + (1-p_i)\hat{\alpha}_i)$.

We analze those terms using the concentration lemmas is Appendix A. In particular, we define

$$\mathcal{G} = \left\{\forall i,\ \bar{E}^+_{T,i} \leq 3c\sqrt{2(T-\tau)\log(dT/\delta)} \text{ and } \bar{E}^-_{T,i} \leq 3c\sqrt{2(T-\tau)\log(dT/\delta)}\right\},$$

$$\mathcal{H} = \left\{\begin{array}{ll} \forall i\in[d],\ \forall t > t_{i-1}, & |\bar{\alpha}_i - \hat{\alpha}_{i,t}| \leq \sqrt{2\sigma^2+c^2}\sqrt{2\frac{\log(dT/\delta)+\log(1+4\tau_{i,t})}{\tau_{i,t}+1}}\,; \\ & |\bar{\beta}_i - \hat{\beta}_{i,t}| \leq \sqrt{2\sigma^2+c^2}\sqrt{2\frac{\log(dT/\delta)+\log(1+4\tau_{i,t})}{\tau_{i,t}+1}} \end{array}\right\},$$

$$\mathcal{I} = \left\{\begin{array}{ll} \forall i\in[d], & \hat{E}^+_{T,i} \leq (T-\tau)\Big(1+\frac{\sqrt{2\log(dT/\delta)}}{\sqrt{T-\tau}}\Big)\max\Big(|\bar{\alpha}_i-\hat{\alpha}_i|, |\bar{\beta}_i-\hat{\beta}_i|\Big) \\ & \hat{E}^-_{T,i} \leq (T-\tau)\Big(1+\frac{\sqrt{2\log(dT/\delta)}}{\sqrt{T-\tau}}\Big)\max\Big(|\bar{\alpha}_i-\hat{\alpha}_i|, |\bar{\beta}_i-\hat{\beta}_t|\Big) \end{array}\right\},$$

$$\mathcal{J} = \left\{\begin{array}{ll} \forall i\in[d], & L^+_{T,i} \leq \frac{3c}{\sqrt{2}}\sqrt{(T-\tau)\log(dT/\delta)}\,, \\ & L^-_{T,i} \leq \frac{3c}{\sqrt{2}}\sqrt{(T-\tau)\log(dT/\delta)} \end{array}\right\}.$$

Applying Lemma 1 and a union bound yields that $\mathbb{P}(\mathcal{G}^c \cup \mathcal{I}^c \cup \mathcal{J}^c) \leq \frac{6\delta}{T}$. Likewise Lemma 2 yields $\mathbb{P}(\mathcal{H}^c) \leq \frac{4\delta}{T}$.

Besides , $\mathcal{G} \cap \mathcal{H} \cap \mathcal{I} \cap \mathcal{J} \subseteq \mathcal{E}$ (calculations assuming $d \geq 2$ and $\tau \leq T/2$, $\log(dt/\delta) \geq 0$ and $\delta \leq dT$). ■

C.2. Proof for the duration of the exploration phase

The following lemma is just a consequence of the definition, but it is particularly useful when analyzing double-greedy approaches, as it limits the range of possible marginal gains to consider when adding/removing items.

Lemma 3 *Let $\mathcal{D}$ be a finite set and f be a submodular set-function. Let $A \subset B \subseteq \mathcal{D}$ and an item $i \in (B \setminus A)$. Then,*

$$\Big(f(A \cup \{i\}) - f(A)\Big) + \Big(f(B \setminus \{i\}) - f(B)\Big) \geq 0\,.$$

Proposition 3 *For each items i, under event $\mathcal{E}$ defined in Proposition 2, `UpdExp` finds a weight p_i such that $l(\hat{\alpha}_{i,t}, \hat{\beta}_{i,t}, p_i) + g_{T,\delta}/\sqrt{\tau_{i,t}} \leq 0$ before $\tau_{i,t}$ has reached $(g_{T,\delta} + \gamma_{T,\delta})^2\, h_{f,i}$.*

Proof We need to look for conditions for Eq. (14)

$$p(\bar{\beta}_i - 3\bar{\alpha}_i) \leq -\frac{g_i + \gamma_{T,\delta}}{\sqrt{\tau_i}} + (\beta_i - 2\bar{\alpha}_i), \qquad p(3\bar{\beta}_i - \bar{\alpha}_i) \leq -\frac{g_i + \gamma_{T,\delta}}{\sqrt{\tau_i}} + \bar{\beta}_i\,, \tag{14}$$

where $\gamma_{T,\delta} = 3\sqrt{(2\sigma^2 + c^2)(\log(dT/\delta) + \log(1+T))}$ to be satisfied.

We consider the different configurations of $(\alpha,\ \beta)$ possible using Lemma 3, which gives 5 zones, and sufficient conditions for the existence of a $p_i \in [0,1]$ satisfying Eq. (14).

	Zone	Threshold of $\frac{\tau_i}{(g_{T,\delta}+\gamma_{T,\delta})^2}$
①	$\bar{\alpha}_i \leq 0,\ \bar{\beta}_i > 0$	$1/\bar{\beta}_i^2$
②	$0 \leq \bar{\alpha}_i \leq \bar{\beta}_i/3$	$1/(\bar{\beta}_i - 2\bar{\alpha}_i)^2$
③	$0 \leq \bar{\beta}_i/3 \leq \bar{\alpha}_i \leq 3\bar{\beta}_i$	$(\bar{\alpha}_i + \bar{\beta}_i)^2/(\bar{\beta}_i - \bar{\alpha}_i)^4$
④	$0 \leq 3\bar{\beta}_i \leq \alpha_i$	$1/(\bar{\alpha}_i - 2\bar{\beta}_i)^2$
⑤	$\bar{\alpha}_i > 0,\ \bar{\beta}_i \leq 0$	$1/\bar{\alpha}_i^2$

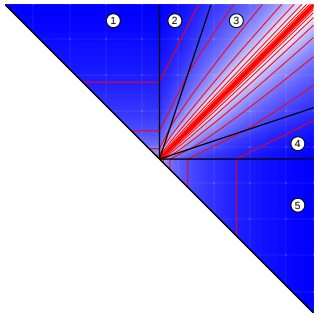


Table 1: Exploration thresholds for `UpdExp`.
The threshold of Table 1 are upper-bounded by the DG-hardness defined in Definition 2.

■