Revisiting Online Submodular Minimization: Gap-Dependent Regret Bounds, Best of Both Worlds and Adversarial Robustness

Shinji Ito [1]

Abstract

In this paper, we consider online decision problems with submodular loss functions. For such problems, existing studies have only dealt with worst-case analysis. This study goes beyond worst-case analysis to show instance-dependent regret bounds. More precisely, for each of the full-information and bandit-feedback settings, we propose an algorithm that achieves a gap-dependent $O(\log T)$-regret bound in the stochastic environment and is comparable to the best existing algorithm in the adversarial environment. The proposed algorithms also work well in the stochastic environment with adversarial corruptions, which is an intermediate setting between the stochastic and adversarial environments.

1. Introduction

This paper considers the *online submodular minimization* problem, a sequential decision-making problem with submodular cost functions. In this problem, a player sequentially chooses a subset $X_t \subseteq [n]$ of a finite set $[n] = \{1, 2, \dots, n\}$ and then gets feedback of the cost function $f_t : 2^{[n]} \to \mathbb{R}$. We suppose that cost functions are *submodular*, i.e., we assume that $f_t(X)+f_t(Y) \geq f_t(X \cup Y)+f_t(X \cap Y)$ holds for any $X, Y \subseteq [n]$. The goal of the player is to minimize the cumulative cost $\sum_{t=1}^{T} f_t(X_t)$ and the performance is evaluated by means of the regret R_T defined as

$$R_T = \max_{X^* \subseteq [n]} \mathbf{E}\left[\sum_{t=1}^{T} f_t(X_t) - \sum_{t=1}^{T} f_t(X^*)\right]. \quad (1)$$

For the feedback information, we consider two different problem settings: the *full-information setting* and the *bandit-feedback setting*. In the former setting, the player gets access to all the information in the cost function f_t, i.e., can observe $f_t(X)$ for any X, after determining X_t. In the latter bandit-feedback setting, the player can observe only $f_t(X_t)$, but the values of $f_t(X)$ for $X \neq X_t$ are not observable.

In many applications of online submodular optimization, it is important to consider a variety of environment models. Submodular functions are closely related to the law of diminishing returns in economics (Bach, 2013; Fujishige, 2005), and therefore, several applications of submodular minimization can be found in the context of marketing. As an application of online submodular minimization, Hazan & Kale (2012) illustrated the problem of maximizing profits by choosing a set of goods to produce. Matsuoka et al. (2021) also pointed out that multi-product price optimization can be formulated in terms of online submodular minimization. In the latter application, for example, the profit can be expressed by a function depending on the set of discounted products, which is supermodular (negation of submodular) under the assumption that the demand of each product is an additive function in prices and products have a relationship of substitute goods (Ito & Fujimaki, 2016). Here, the cost function f_t corresponds to the market response models, which are supposed to be time-varying in nature. These examples motivate us to consider a variety of dynamic environment models that characterize the behavior of f_t.

Previous studies Existing studies on online submodular minimization focus on adversarial models, in which no stochastic models for f_t are assumed, but $\{f_t\}_{t=1}^{T}$ is an *arbitrary* sequence of submodular functions. For this model, Hazan & Kale (2012) have proposed a computationally efficient algorithm that achieves $O(\sqrt{nT})$-regret for the full-information setting. This regret upper bound is *worst-case optimal* up to a constant. In fact, any algorithm suffers regret of at least $\Omega(\sqrt{nT})$ for some (worst-case) input sequences of f_t. For the bandit-feedback setting, they have presented an algorithm with $O(nT^{2/3})$-regret as well. It is not known if this is tight at present.

A drawback of the worst-case optimal algorithms for the adversarial model is that they tend to be too conservative. In the adversarial model, because no assumptions are made

[1]NEC Corporation, Tokyo, Japan. Correspondence to: Shinji Ito <i-shinji@nec.com>.

Proceedings of the 39th International Conference on Machine Learning, Baltimore, Maryland, USA, PMLR 162, 2022.

Table 1. Regret bounds for online submodular minimization with full-information feedback.

Model	Alg. 2 in [HK12][1]	FTL	Alg. 1 in **[This work]**	Lower bound
Adversarial	$O(\sqrt{nT})$	–	$O(\sqrt{nT})$	$\Omega(\sqrt{nT})$
Stochastic	$O(\sqrt{nT})$	$O\left(\frac{n}{\Delta}\right)$	$O\left(\min\left\{\sqrt{nT}, \frac{n}{\Delta}\right\}\right)$	$\Omega\left(\min\left\{\sqrt{nT}, \frac{1}{\Delta}\right\}\right)$
Sto. with Adv.	$O(\sqrt{nT})$	–	$O\left(\min\left\{\sqrt{nT}, \frac{n}{\Delta} + \sqrt{\frac{Cn}{\Delta}}\right\}\right)$	$\Omega\left(\min\left\{\sqrt{nT}, \frac{1}{\Delta} + \sqrt{\frac{C}{\Delta}}\right\}\right)$

Table 2. Regret bounds for online submodular minimization with bandit feedback.

Model	Alg. 3 in [HK12][1]	Alg. 2 in **[This work]**	Lower bound
Adversarial	$O(nT^{2/3})$	$O(nT^{2/3}(\log T)^{1/3})$	$\Omega(n\sqrt{T})$
Stochastic	$O(nT^{2/3})$	$O\left(\min\left\{nT^{2/3}(\log T)^{1/3}, \frac{n^3 \log T}{\Delta^2}\right\}\right)$	$\Omega\left(\min\left\{n\sqrt{T}, \frac{n^2}{\Delta}\right\}\right)$
Sto. with Adv.	$O(nT^{2/3})$	$\tilde{O}\left(\min\left\{nT^{2/3}, \frac{n^3}{\Delta^2} + \left(\frac{C^2 n^3}{\Delta^2}\right)^{1/3}\right\}\right)$	$\Omega\left(\min\left\{n\sqrt{T}, \frac{n}{\Delta} + \sqrt{\frac{Cn}{\Delta}}\right\}\right)$

about the input sequence, *worst-case analysis* is used, i.e., the algorithm is evaluated by its performance on the most *difficult* input sequence $\{f_t\}_t^T$. In practice, however, such difficult environments are not always the case. For example, if the environment can be assumed to be stationary, then the stochastic model is reasonable, in which f_t follows an unknown distribution, i.i.d., for all t. In this case, a simple follow-the-leader (FTL) algorithm achieves $O(\frac{n}{\Delta})$-regret, where $\Delta > 0$ represents the *suboptimality gap* parameter defined in Section 3, as can be concluded from the standard analysis for FTL, e.g., from Theorem 1 by Degenne & Perchet (2016). This bound is independent of the number T of rounds, and better than the worst-case optimal bound of $O(\sqrt{nT})$ for sufficiently large T.

Motivations In order to overcome the shortcoming of algorithms for the adversarial model, this study aims to go beyond worst-case analysis by developing algorithms that adapt to the tendencies of the environment. In particular, we focus on *best-of-both-worlds* (BOBW) algorithms that perform well for both stochastic and adversarial models. Furthermore, we consider the *stochastic model with adversarial corruptions*, which is an intermediate setting between stochastic and adversarial models. In this setting, the adversary corrupts the stochastically generated losses, and the total amount of disturbance is represented by the *corruption level parameter* C, of which definition is given in Section 3. If $C = 0$, this model is equivalent to the stochastic model. By way of contrast, if C is unconstrained (e.g., $C = O(T)$), the model is equivalent to the adversarial model. From these facts, we can see that the stochastic model with adversarial corruptions is a comprehensive setting that includes both stochastic and adversarial models.

[1] [HK12] stands for a reference to the paper by Hazan & Kale (2012).

Contributions The contribution of this work is to develop BOBW algorithms for online submodular minimization. The proposed algorithm (Algorithm 1) for the full-information setting achieves $O(\sqrt{nT})$-regret in the adversarial model and $O(\frac{n}{\Delta})$-regret in the stochastic model. In addition, for the stochastic model with adversarial corruption, Algorithm 1 has the regret bound of $O(\frac{n}{\Delta} + \sqrt{\frac{Cn}{\Delta}})$. This can be achieved without any prior knowledge on the corruption level C. For the bandit-feedback setting, we propose an algorithm (Algorithm 2) that achieves regret bounds of $O(nT^{2/3}(\log T)^{1/3})$ in the adversarial model, of $O(\frac{n^3 \log T}{\Delta^2})$ in the stochastic model, and of $O(\frac{n^3 \log T}{\Delta^2} + (\frac{C^2 n^3 \log T}{\Delta^2})^{1/3})$ in the stochastic model with adversarial corruptions. The regret bounds for full-information and bandit-feedback settings are summarized in Tables 1 and 2, respectively.

In the design of algorithms, we combine two main technical elements, the *Lovász extension* (Lovász, 1983; Fujishige, 2005) of submodular functions, and the framework of *follow-the-regularized-leader* (Cesa-Bianchi & Lugosi, 2006) with adaptive learning rates. The Lovász extension is a technique that extends the domain of set functions to continuous regions, by which we can reduce submodular minimization into convex optimization problems. Such a technique has been employed in the work by Hazan & Kale (2012), as well as in several studies on submodular function minimization (Ito, 2019; Chakrabarty et al., 2017; Axelrod et al., 2020; Matsuoka et al., 2021).

A key technique in this paper to go beyond the existing studies is to use the follow-the-regularized-leader method with time-varying regularizers, instead of the simple subgradient descent method with constant learning rate employed, e.g., by Hazan & Kale (2012). By designing regularizers in a sophisticated way, we obtain regret bounds depending on

the output sequence. More concretely, this paper provides a novel regularizer and an update rule of learning rates for each of full-information and bandit-feedback settings, so that we can obtain regret bounds depending on the output sequences. Such output-dependent regret bounds lead to BOBW properties and robustness to adversarial corruptions, via the so-called self-bounding technique (Wei & Luo, 2018; Zimmert & Seldin, 2021; Masoudian & Seldin, 2021). A further description and references of the self-bounding technique are provided in the next section.

2. Related Work

Submodular function minimization (Iwata, 2008; McCormick, 2005) has been studied for a long time, with a wide range of applications such as combinatorial optimization problems (Fujishige, 2005), image segmentation (Jegelka & Bilmes, 2011b; Kohli & Torr, 2010), supervised learning with structured regularization (Bach, 2013), training data subset selection (Lin & Bilmes, 2010), and multi-product price optimization (Ito & Fujimaki, 2016). For exact optimization, after the first polynomial-time algorithm was proposed by Grötschel et al. (1981), many improvements followed (Iwata et al., 2001; Schrijver, 2000; Orlin, 2009; Iwata, 2003; Lee et al., 2015). For approximate optimization, Chakrabarty et al. (2017) and Axelrod et al. (2020) have proposed subgradient-descent-based algorithms that find an ε-additive approximate minimizer run in $\tilde{O}(n^{5/3}/\varepsilon^2)$-time and in $\tilde{O}(n/\varepsilon^2)$-time, respectively. Ito (2019) proposed algorithms for submodular function minimization with noisy function value oracles, which can be considered as a special case of online submodular minimization (Hazan & Kale, 2012). Matsuoka et al. (2021) also considered online submodular minimization problems and proposed algorithms with tracking-regret bounds. Jegelka & Bilmes (2011a) have studied online constrained submodular minimization problems over some combinatorial structures and proposed online approximation algorithms for them.

Best-of-both-worlds (BOBW) algorithms have been studied extensively for a variety of online decision problems, including the problem of prediction with expert advice (Gaillard et al., 2014; Luo & Schapire, 2015; De Rooij et al., 2014; Mourtada & Gaïffas, 2019), the multi-armed bandit problem (Bubeck & Slivkins, 2012; Seldin & Slivkins, 2014; Auer & Chiang, 2016; Zimmert & Seldin, 2021; Ito, 2021c; Wei & Luo, 2018), combinatorial semi-bandit problems (Zimmert & Seldin, 2019; Ito, 2021a), online learning with feedback graphs (Erez & Koren, 2021), linear bandit problems (Lee et al., 2021), and episodic MDPs (Jin et al., 2021; Jin & Luo, 2020). Among these, studies by Gaillard et al. (2014) and Luo & Schapire (2015) are particularly relevant to this study. These studies show the BOBW property via regret bounds depending on the output sequence, using the self-bounding technique. The self-bounding technique, which is employed in several studies on BOBW algorithms (Gaillard et al., 2014; Luo & Schapire, 2015; Wei & Luo, 2018; Zimmert & Seldin, 2019; 2021; Ito, 2021c; Erez & Koren, 2021), has been used to show the robustness to adversarial corruption as well, e.g., for the multi-armed bandit (Zimmert & Seldin, 2021; Ito, 2021c; Masoudian & Seldin, 2021) and for online learning with feedback graphs (Erez & Koren, 2021).

For adversarial corruption in the stochastic environments, several different models have been studied, including the *oblivious corruption model* (Lykouris et al., 2018; Gupta et al., 2019; Bogunovic et al., 2020) and the *targeted corruption model* (Jun et al., 2018; Hajiesmaili et al., 2020; Liu & Shroff, 2019; Bogunovic et al., 2021; Amir et al., 2020). These two models differ in the information that can be accessed by adversaries. Specifically, in the former oblivious corruption model, the adversary corrupts the cost function f_t before observing the player's action X_t, whereas, in the latter targeted corruption model, the corruption on f_t is determined depending on the player's action X_t. This study deals with the former model. It is known that some BOBW algorithms based on the self-bounding technique work well for the oblivious corruption model (Zimmert & Seldin, 2021; Erez & Koren, 2021).

In the context of online submodular optimization, there is more research on maximization than on minimization. As typical submodular function maximization problems are computationally hard, unlike minimization problems, approximate regrets are used as evaluation metrics. For online size-constrained monotone submodular maximization, Streeter & Golovin (2008) provided algorithms achieving sublinear $(1 - 1/\mathrm{e})$-approximate regret. Similarly for online unconstrained non-monotone submodular maximization, Roughgarden & Wang (2018) have proposed an algorithm with a $(1/2)$-approximate-regret bound. These regret bounds have been improved by Harvey et al. (2020). Besides these, the study on online submodular maximization has been extended to include k-submodular maximization (Soma, 2019) and continuous submodular maximization (Chen et al., 2018a;b; Zhang et al., 2019). Additional topics and applications of submodular functions can be found in the paper by Bilmes (2022).

3. Problem Setting

A player is given $T \geq 2$ and $n \geq 1$, which represent the number of rounds and the size of the underlying set associated with cost functions, respectively. In each round $t \in [T]$, the player (randomly) chooses a subset $X_t \in [n]$ of the underlying set $[n] := \{1, 2, \cdots, n\}$. We consider two different settings w.r.t. the feedback information about the cost function. In the *full-information* setting, after choosing X_t, the player can observe all function values of the cost

function $f_t : 2^{[n]} \to [0,1]$, i.e., can observe $f_t(X)$ for any $X \in [n]$ after choosing X_t. In the *bandit-feedback* setting, the player can observe only the value $f_t(X_t)$ for the chosen action X_t. Cost functions f_t are here assumed to be submodular, i.e.,

$$f_t(X \cap Y) + f_t(X \cup Y) \leq f_t(X) + f_t(Y) \qquad (2)$$

holds for any $X, Y \subseteq [n]$. In each round t, the environment decides f_t before the player chooses X_t.

The performance of the algorithm is evaluated in terms of the (pseudo-) regret R_T defined by (1) which is the difference between the cumulative loss for the algorithm's choice $\{X_t\}_{t=1}^T$ and that for an optimal fixed action X^* that minimizes the cumulative loss in expectation.

This work considers three different models for the environment: the *stochastic model*, the *adversarial model*, and *the stochastic model with adversarial corruptions*.

Stochastic model In the stochastic model, we assume there exists an unknown distribution $\mathcal{D}$ over a set of submodular functions, and $f_t : 2^{[n]} \to [0,1]$ follows $\mathcal{D}$ i.i.d. for all $t = 1, 2, \ldots, T$. Denote

$$\bar{f}(X) = \mathop{\mathbf{E}}_{f \sim \mathcal{D}}[f(X)].$$

We define $X^* \subseteq [n]$ and the *suboptimality gap* $\Delta \geq 0$ by

$$X^* \in \mathop{\arg\min}_{X \subseteq [n]} \bar{f}(X), \quad \Delta = \min_{X \in 2^{[n]} \setminus \{X^*\}} (\bar{f}(X) - \bar{f}(X^*)).$$

We note that no prior knowledge on the parameter Δ is given to the player. When considering the gap-dependent regret bounds, we assume $\Delta > 0$, i.e., we assume that the minimizer of $\bar{f}$ is *unique*. Similar assumptions were also made in previous works (Gaillard et al., 2014; Luo & Schapire, 2015; Wei & Luo, 2018; Mourtada & Gaïffas, 2019). We note that, in this model, the regret defined by (1) can be rewritten as $R_T = \sum_{t=1}^T (\mathbf{E}\left[\bar{f}(X_t)\right] - \bar{f}(X^*))$ with $X^* \in \arg\min_{X \subseteq [n]} \bar{f}(X)$, which follows from the fact that X_t is dependent of f_t.

Adversarial model In the adversarial model, the cost function may behave in a non-stationary manner. More precisely, in this model, the environment chooses submodular functions $f_t : 2^{[n]} \to [0,1]$ depending on the output sequence $(X_1, X_2, \ldots, X_{t-1})$ which the algorithm has chosen so far. As is obvious, this adversarial model includes the stochastic model as a special case.

Stochastic model with adversarial corruptions In the stochastic model with adversarial corruptions (SwA model), a *temporary cost* $f'_t : 2^{[n]} \to [0,1]$ is chosen from an unknown stationary distribution $\mathcal{D}$, and then the adversary *corrupts* f'_t to produce the cost $f_t : 2^{[n]} \to [0,1]$. We here assume that both f_t and f'_t are submodular. We define the *corruption level* C by

$$C = \sum_{t=1}^T \max_{X \subseteq [n]} |\, \mathbf{E}\left[f_t(X) - f'_t(X)\right] |, \qquad (3)$$

which measures the total amount of corruption by the adversary. From the definition, the corruption level is bounded as $0 \leq C \leq T$. This study deals with a problem setup in which the player is given no prior knowledge on the corruption level C. Note that the player can only observe information about f_t, but not about f'_t. Also, the regret is defined on the basis of f, not on f'_t. In the SwA model, we define the parameter Δ in the same way as in the stochastic model, from the distribution $\mathcal{D}$ that generates f'_t.

When considering SwA model, this study investigates the regret R_T defined in terms of cost functions f_t *after* corruption, as in (1). We note that some existing studies focus on an alternative regret notion R'_T defined in terms of cost functions f'_t *before* corruption. When comparing R_T and R'_T, which is more natural depends on the application scenario. In the application to price optimization (described in Section 1), if the corruption is due to seasonal demand fluctuations and the actual profit is corrupted, using R_T is probably the right choice as it reflects the actual total profit. If the corruption is due to the miscalculation in accounting and the actual profit is not corrupted, its natural to evaluate the performance w.r.t. R'_T. A discussion on the difference between R_T and R'_T can be found in Section 5.2 of (Gupta et al., 2019). One of the reasons we chose R_T is because a bound on R_T implies a bound on R'_T. In fact, as we obtain $|R_T - R'_T| = O(C)$ from the definition, a regret bound of $R_T = O(\mathcal{R} + \sqrt{C\mathcal{R}})$ implies that $R'_T = O(\mathcal{R} + \sqrt{C\mathcal{R}} + C) = O(\mathcal{R} + C)$ where we used the AMGM inequality. On the other hand, $R'_T = O(\mathcal{R} + C)$ does not imply $R_T = O(\mathcal{R} + \sqrt{C\mathcal{R}})$.

4. Proposed Algorithms

4.1. Preliminary

Lovász extension Given a set function $f : 2^{[n]} \to \mathbb{R}$, we can define the *Lovász extension* $\tilde{f} : [0,1]^n \to \mathbb{R}$ as follows: For $x = (x_1, x_2, \ldots, x_n)^\top \in [0,1]^n$ and $u \in [0,1]$, let $H_u(x)$ denote the set of indices i for which $x_i \geq u$, i.e., $H_u(x) := \{i \in [n] \mid x_i \geq u\}$. Then the Lovász extension $\tilde{f}$ is defined by

$$\tilde{f}(x) = \mathop{\mathbf{E}}_{u \sim \mathrm{Unif}([0,1])} \left[f(H_u(x))\right], \qquad (4)$$

where $\mathrm{Unif}([0,1])$ means a uniform distribution over $[0,1]$. The Lovász extension $\tilde{f}$ is convex if and only if f is submodular (Lovász, 1983). From this definition, for any

$x \in [0,1]^n$, and for any permutation $\sigma : [n] \to [n]$ such that $x_{\sigma(i)} \leq x_{\sigma(i+1)}$ for any $i \in [n-1]$, $\tilde{f}(x)$ can be expressed as

$$\tilde{f}(x) = \sum_{i=0}^{n} (x_{\sigma(i+1)} - x_{\sigma(i)}) f(\sigma([i])), \tag{5}$$

where we denote $\sigma([i]) = \{\sigma(j) \mid j \in [i]\}$ and exceptionally define $x_{\sigma(0)} = 0$ and $x_{\sigma(n+1)} = 1$. Hence, $h(\sigma) \in \mathbb{R}^n$ defined in the following is a subgradient of $\tilde{f}$ at x:

$$h(\sigma) = \sum_{i=0}^{n} f(\sigma([i]))\rho_i(\sigma), \tag{6}$$

$$\rho_i(\sigma) = \begin{cases} \chi_{\sigma(1)} & i = 0 \\ \chi_{\sigma(i+1)} - \chi_{\sigma(i)} & i \in [n-1] \\ -\chi_{\sigma(n)} & i = n \end{cases}, \tag{7}$$

where $\chi_i \in \{0,1\}^n$ expresses the indicator vector of i, i.e., $\chi_{ij} = 1$ if and only if $i = j$. From the definition of subgradients, we have $\tilde{f}(y) - \tilde{f}(x) \geq \langle h(\sigma), y - x \rangle$ for any $y \in [0,1]^n$. In the regret analysis for the proposed algorithms, we use the following lemma:

Lemma 4.1. *If $f : 2^{[n]} \to [0,1]$ is a submodular function, subgradients $h(\sigma)$ of the Lovász extension of f defined in (6) are bounded as $\|h(\sigma)\|_1 \leq 4$ for any permutation σ.*

The proof of this lemma can be found, e.g., in the paper by Hazan & Kale (2012); see the proof of their Lemma 8.

Follow the regularized leader The follow-the-regularized-leader (FTRL) method is a generic approach for online convex optimization. Let $\Omega \subseteq \mathbb{R}^n$ denote a convex feasible region. The update rule of FTRL can be expressed as follows:

$$x_t \in \arg\min_{x \in \Omega} \left\{ \left\langle x, \sum_{s=1}^{t-1} g_s \right\rangle + \psi_t(x) \right\}, \tag{8}$$

where g_t denotes a subgradient of the cost function f_t at x_t and ψ_t is a regularizer that is a smooth convex function over Ω. Regret bounds for FTRL can be analyzed using the Bregman divergences. Let D_t denote the Bregman divergence associated with ψ_t:

$$D_t(x,y) = \psi_t(x) - \psi_t(y) - \langle \nabla \psi_t(y), x - y \rangle, \tag{9}$$

where $\nabla \psi_t$ represents the gradient of ψ_t.

For the FTRL method defined by (8), we have the following regret bound:

Lemma 4.2. *For $x_t \in \Omega$ defined by (8) and for any $x^* \in \Omega$, we have*

$$\sum_{t=1}^{T} (f_t(x_t) - f_t(x^*)) \leq \sum_{t=1}^{T} (\langle g_t, x_t - x_{t+1} \rangle - D_t(x_{t+1}, x_t)) + \sum_{t=1}^{T} (\psi_t(x_{t+1}) - \psi_{t+1}(x_{t+1})) + \psi_{T+1}(x^*) - \psi_1(x_1).$$

For the proof of this lemma, see, e.g., Exercise 28.12 of the book by Lattimore & Szepesvári (2020).

4.2. Algorithm for the full-information setting

This subsection provides an algorithm for the full-information setting. The proposed algorithm is based on FTRL (8) with regularizers defined as

$$\psi_t(x) = \sum_{i=1}^{n} \lambda_{ti} \phi(x_i), \quad \phi(z) = z \log z + (1-z)\log(1-z), \tag{10}$$

where $\lambda_{ti} > 0$ corresponds to learning rates, which are defined in (11) below. This regularization term makes a significant difference from the existing study by Hazan & Kale (2012). The function ϕ corresponds to the *negative entropy*, which is used to construct BOBW algorithms for the problem of prediction with expert advice (Gaillard et al., 2014; Luo & Schapire, 2015). Intuitively, when we use the regularization terms in (10), the closer x is to the boundary of $[0,1]^n$, the stronger the regularization effect. As a result, we get an $\{x_t\}$-dependent regret upper bound as shown in Proposition 4.5 below, which becomes smaller as x_t's get closer to the boundary. This allows us to prove the BOBW property and robustness to adversarial corruption.

The proposed algorithm computes x_t on the basis of follows-the-regularized-leader method defined in (8) with the regularizer function ψ_t defined in (10) and with subgradients g_s of the Lovász extensions $\tilde{f}_s$ of the cost function f_s. After computing x_t, the algorithm output $X_t = H_u(x_t) = \{i \in [n] \mid x_{ti} \geq u_t\}$, where u is chosen from a uniform distribution over $[0,1]$. We then have $\mathbf{E}[f_t(X_t)] = \mathbf{E}[\tilde{f}(x_t)]$ from (4).

Let us show how we can compute x_t defined with (8) and (10). As we have $\frac{\mathrm{d}\phi(z)}{\mathrm{d}z} = \log \frac{z}{1-z}$, the vector $x_t \in [0,1]^n$ determined by the FTRL method (8) can be expressed as

$$x_{ti} = \frac{1}{1 + \exp(G_{ti}/\lambda_{ti})}, \quad G_t = \sum_{s=1}^{t-1} g_s = \sum_{s=1}^{t-1} h_s(\sigma_s),$$

where $\sigma_s : [n] \to [n]$ is a permutation over $[n]$ such that $x_{s\sigma_s(i)} \leq x_{s\sigma_s(i+1)}$ for $i \in [n-1]$ and $h_s(\sigma_s) \in \mathbb{R}^n$ is define as in (6) with $f = f_s$ and $\sigma = \sigma_s$.

We set the learning rate λ_{ti} by

$$\lambda_{ti} = 2 + \frac{1}{\log 2} \sum_{s=1}^{t-1} \lambda_{si} \cdot \nu\left(\frac{g_{si}}{\lambda_{si}}, x_{si}\right) \tag{11}$$

where ν is defined as

$$\nu(g,z) = \log(1 - z + z\exp(g)) - zg \tag{12}$$

Algorithm 1 An algorithm for the full-information setting

Require: The size n of the underlying set.

1: Initialize G_{ti} by $G_{1i} = 0$ for all $i \in [n]$.
2: **for** $t = 1, 2, \dots$ **do**
3: Compute $x_t \in [0,1]^n$ by $x_{ti} = \frac{1}{1+\exp(G_{ti}/\lambda_{ti})}$ for each $i \in [n]$, where λ_{ti} is defined in (11).
4: Let $\sigma_t : [n] \to [n]$ be a permutation over $[n]$ such that $x_{t\sigma_t(i)} \le x_{t\sigma_t(i+1)}$ for all $i \in [n-1]$.
5: Pick u_t from a uniform distribution over $[0,1]$.
6: Output $X_t = H_{u_t}(x_t) = \{i \in [n] \mid x_{ti} \ge u_t\}$.
7: Compute $g_t \in \mathbb{R}^d$ defined as

$$g_t = h_t(\sigma_t) = \sum_{i=0}^{n} f_t(\sigma_t([i]))\rho_i(\sigma_t), \tag{13}$$

where $\rho_i(\sigma_t)$ is defined in (6).
8: Update G_t by $G_{t+1} = G_t + g_t$.
9: **end for**

for all $z \in [0,1]$ and $g \in \mathbb{R}$.

The proposed algorithm for the full-information setting can be summarized in Algroithm 1. The computational time per round required by Algorithm 1 is $O(n(\mathrm{EO}\log n))$, where EO denotes the time taken to evaluate $f_t(X)$ for a single input X. In fact, computing σ_t can be done in $O(n \log n)$, each λ_{ti} can be computed in $O(1)$ time via the updated rule of $\lambda_{t+1,i} = \lambda_{ti} + \frac{1}{\log 2}\lambda_{ti}\nu(\frac{g_{ti}}{\lambda_{ti}}, x_{ti})$, and the other computation can be done in $O(n)$ time.

In the rest of this subsection, we show the following regret bound:

Theorem 4.3. *The regret for Algorithm 1 is bounded as*

$$R_T = \begin{cases} O\left(\sqrt{nT}\right) & \textit{(Adv. model)} \\ O\left(\frac{n}{\Delta}\right) & \textit{(Sto. model)} \\ O\left(\frac{n}{\Delta} + \sqrt{\frac{Cn}{\Delta}}\right) & \textit{(SwA model)} \end{cases}. \tag{14}$$

To show this bound, we use Lemma 4.2. When ψ_t is defined as (10), the parts of $\langle g_t, x_t - x_{t+1}\rangle - D_t(x_{t+1}, x_t)$ can be bounded as follows:

Lemma 4.4. *If ψ_t is defined by (10) with $\lambda_{ti} \ge 2$, we have*

$$\begin{aligned} &\langle g_t, x_t - x_{t+1}\rangle - D_t(x_{t+1}, x_t) \\ &\le \sum_{i=1}^{n} \lambda_{ti} \cdot \nu\left(\frac{g_{ti}}{\lambda_{ti}}, x_{ti}\right) \le \sum_{i=1}^{n} \frac{g_{ti}^2 \min\{x_{ti}, 1 - x_{ti}\}}{\lambda_{ti}}. \end{aligned}$$

All omitted proof can be found in the appendix. Combining this lemma with Lemmas 4.1 and 4.2, and from the definition (11) of the learning rates, we obtain the following proposition:

Proposition 4.5. *The regret for Algorithm 1 is bounded as follows:*

$$R_T = O\left(\mathbf{E}\left[\sqrt{n\sum_{t=1}^{T}\max_{i\in[n]}\{\min\{x_{ti}, 1-x_{ti}\}\}}\right] + n\right).$$

As we have $\max_{i\in[n]}\{\min\{x_{ti}, 1-x_{ti}\}\} \le 1$, the regret bound of $O(\sqrt{nT})$ in Theorem 4.3 immediately follows from Proposition 4.5. The other regret bounds in Theorem 4.3 can be shown by using the self-bounding technique and the following lemma.

Lemma 4.6. *In the SwA model, we have*

$$\sum_{t=1}^{T}\mathbf{E}[f_t(X_t) - f_t(X^*)] \ge \Delta\sum_{t=1}^{T}\mathrm{Prob}[X_t \ne X^*] - 2C$$

for any $X^* \subseteq [n]$. This inequality implies

$$R_T \ge \Delta\,\mathbf{E}\left[\sum_{t=1}^{T}\max_{i\in[n]}\{\min\{x_{ti}, 1-x_{ti}\}\}\right] - 2C. \tag{15}$$

We are now ready to prove Theorem 4.3.

Proof of Theorem 4.3. The regret bound of $O(\sqrt{nT})$ can be shown immediately from Proposition 4.5. As we have $\max_{i\in[n]}\{\min\{x_{ti}, 1-x_{ti}\}\} \le 1$, from Proposition 4.5, we have $R_T = O(\sqrt{nT}+n)$. Further, from the assumption that $|f_t(X)| = O(1)$ for all X, we have a trivial regret bound of $R_T = O(T)$. Combining these two regret upper bounds, we obtain $R_T = O(\min\{\sqrt{nT}+n, T\}) = O(\sqrt{nT})$. Indeed, in the case of $T \le n$, we have $\min\{\sqrt{nT}+n, T\} \le T \le \sqrt{nT}$, and in the other case of $T > n$, we have $\min\{\sqrt{nT}+n, T\} \le \sqrt{nT}+n \le 2\sqrt{nT}$.

Improved regret bounds for the stochastic model (with adversarial corruptions) can be shown from Proposition 4.5 and Lemma 4.6. Denote $Q = \mathbf{E}\left[\sum_{t=1}^{T}\max_{i\in[n]}\{\min\{x_{ti}, 1-x_{ti}\}\}\right]$. Then Proposition 4.5 and Jensen's inequality implies that $R_T \le B(\sqrt{nQ}+n)$ for some constant $B > 0$, and (15) can be rewritten as $R_T \ge \Delta Q - 2C$. By combining these two inequalities, for any $\eta > 0$, we obtain

$$\begin{aligned} R_T &= (1+\eta)R_T - \eta R_T \\ &\le (1+\eta)B\left(\sqrt{nQ}+n\right) - \eta(\Delta Q - 2C) \\ &= (1+\eta)B\sqrt{nQ} - \eta\Delta Q + \eta(Bn+2C) + Bn \\ &\le \frac{(1+\eta)^2B^2n}{4\eta\Delta} + \eta(Bn+2C) + Bn \\ &= \frac{B^2n}{4\eta\Delta} + \eta\left(\frac{B^2n}{4\Delta} + Bn + 2C\right) + Bn + \frac{B^2n}{2\Delta}, \end{aligned}$$

where the second inequality follows from the fact that $ax - bx^2 = -b(x - \frac{a}{2b})^2 + \frac{a^2}{4b} \leq \frac{a^2}{4b}$ holds for any $a \geq 0, b > 0$ and $x \in \mathbb{R}$. We applied this inequality with $a = (1+\eta)B\sqrt{n}$, $b = \eta\Delta$ and $x = \sqrt{Q}$. By setting[2] $\eta = \left(1 + \frac{4\Delta(Bn+2C)}{B^2 n}\right)^{-1/2}$, we obtain

$$\begin{aligned} R_T &\leq 2\sqrt{\frac{B^2 n}{4\Delta}\left(\frac{B^2 n}{4\Delta} + Bn + 2C\right)} + Bn + \frac{B^2 n}{2\Delta} \\ &\leq \frac{B^2 n}{\Delta} + \sqrt{\frac{2B^2 C}{\Delta}} + \sqrt{\frac{B^3 n^2}{\Delta}} + Bn \\ &= O\left(\frac{n}{\Delta} + \sqrt{\frac{Cn}{\Delta}}\right), \end{aligned}$$

where the second inequality follows from $\sqrt{x+y} \leq \sqrt{x} + \sqrt{y}$ that holds for any $x, y \geq 0$, and the last inequality follows from $B = O(1)$ and $\Delta \leq 1$. This completes the proof for the regret bounds for the stochastic model and for the stochastic model with adversarial corruptions in Theorem 4.3. □

4.3. Algorithm for the bandit-feedback setting

This subsection provides an algorithm for the bandit-feedback setting. The proposed algorithm compute x_t and σ_t in a similar way as in Section 4.2. After computing x_t, we choose X_t in a similar way to Algorithm 3 in the paper by Hazan & Kale (2012). We set $X_t = \sigma_t([i_t])$, where $i_t \in \{0, 1, \ldots, n\}$ is chosen so that

$$\begin{aligned} &\mathrm{Prob}[i_t = i | x_t] = p_t(i) \\ &= (1-\gamma_t)(x_{t\sigma_t(i+1)} - x_{t\sigma_t(i)}) + \frac{\gamma_t}{n+1}, \end{aligned} \tag{16}$$

where $\gamma_t \in [0,1]$ is the *exploration rate* parameter defined in (22) below. We then have

$$\mathbf{E}[f_t(X_t)|x_t] \leq (1-\gamma_t)\tilde{f}_t(x_t) + \gamma_t \leq \tilde{f}_t(x_t) + 2\gamma_t, \tag{17}$$

as provided in Lemma 15 of the paper (Hazan & Kale, 2012). After observing $f_t(X_t)$, we compute $\hat{g}_t \in \mathbb{R}^n$ defined as

$$\hat{g}_t = \frac{1}{p_t(i_t)} f_t(X_t)\rho_{i_t}(\sigma_t). \tag{18}$$

This vector $\hat{g}_t$ is an unbiased estimator of a subgradient $g_t = h_t(\sigma_t)$ of $\tilde{f}_t(x)$ at x_t. For the bandit setting, we use the regularization functions ψ_t defined as follows:

$$\psi_t(x) = -\lambda_t \sum_{i=1}^{n} \left(\log(x_i) + \log(1-x_i)\right). \tag{19}$$

[2]The parameter η does not appear in the algorithm but appears only in the analysis. Therefore, we do not need C as an input.

Algorithm 2 An algorithm for the bandit-feedback setting

Require: The size n of the underlying set, and the number T of rounds.
1: Initialize $\hat{G}_{ti}$ by $\hat{G}_{1i} = 0$ for all $i \in [n]$.
2: **for** $t = 1, 2, \ldots, T$ **do**
3: Compute $x_t \in [0,1]^n$ by $x_{ti} = \zeta(\hat{G}_{ti}/\lambda_t)$ for each $i \in [n]$, where ζ and λ_t are defined in (20) and (22), respectively.
4: Let $\sigma_t : [n] \to [n]$ be a permutation over $[n]$ such that $x_{t\sigma_t(i)} \leq x_{t\sigma_t(i+1)}$ for all $i \in [n-1]$.
5: Pick $i_t \in \{0, 1, \ldots, n\}$ with the probability defined in (16), where γ_t is defined in (22).
6: Output $X_t = \sigma_t([i_t]) = \{\sigma_t(j) \mid j \in [i_t]\}$, and get feedback of $f_t(X_t)$.
7: Compute $\hat{g}_t \in \mathbb{R}^d$ defined in (18), where $\rho_i(\sigma_t)$ is given in (6).
8: Update $\hat{G}_t$ by $\hat{G}_{t+1} = \hat{G}_t + \hat{g}_t$.
9: **end for**

Then the vector x_t given by the FTRL method (8) with $g_t = \hat{g}_t$ can be expressed as $x_{ti} = \zeta(\frac{\hat{G}_{ti}}{\lambda_t})$, where $\hat{G}_t = \sum_{s=1}^{t-1} g_s$ and $\zeta(z)$ is defined by

$$\zeta(z) = \begin{cases} \frac{1}{2}\left(1 + \frac{2}{g} - \sqrt{1 + \frac{4}{g^2}}\right) & (g > 0) \\ \frac{1}{2}\left(1 + \frac{2}{g} + \sqrt{1 + \frac{4}{g^2}}\right) & (g < 0) \\ 1/2 & (g = 0) \end{cases}. \tag{20}$$

Define a vector $v_t = (v_{ti})_{i=1}^n \in [0,1]^n$ by

$$v_{ti} = \min\{x_{ti}, 1 - x_{ti}\}. \tag{21}$$

We set the learning rate λ_t and the exploration rate γ_t as follows:

$$\lambda_t = 2n + \left(\frac{1}{\sqrt{n}\log T}\sum_{s=1}^{t-1}\|v_s\|_2\right)^{2/3}, \quad \gamma_t = \sqrt{\frac{n}{\lambda_t}}\|v_t\|_2. \tag{22}$$

The proposed algorithm for the bandit setting is summarized in Algorithm 2. The computational time per round required by Algorithm 2 is $O(n \log n + \mathrm{EO})$, where EO denotes the time taken to evaluate $f_t(X)$ for a single input X.

Algorithm 2 enjoys the following regret bound:

Theorem 4.7. *The regret for Algorithm 2 is bounded as*

$$R_T = \begin{cases} O\left(nT^{2/3}(\log T)^{1/3}\right) & \textit{(Adv. model)} \\ O\left(\frac{n^3 \log T}{\Delta^2}\right) & \textit{(Sto. model)} \\ O\left(\frac{n^3 \log T}{\Delta^2} + \left(\frac{C^2 n^3 \log T}{\Delta^2}\right)^{1/3}\right) & \textit{(SwA model)} \end{cases}.$$

The proof of this theorem starts with the following lemma:

Lemma 4.8. *If X_t and $\hat{g}_t$ are given with (16) and (18), the regret can be bounded as*

$$R_T \leq \mathbf{E}\left[\sum_{t=1}^{T} \langle \hat{g}_t, x_t - x^* \rangle + 2\sum_{t=1}^{T} \gamma_t\right] + 4, \quad (23)$$

where the vector $x^* = (x_i^*)_{i=1}^n = [\frac{1}{T}, 1-\frac{1}{T}]^n$ is defined as $x_i^* = (1-\frac{2}{T})\mathbf{1}[i \in X^*] + \frac{1}{T}$.

This can be shown from (17) and the fact that $\hat{g}_t$ is an unbiased estimator of a subgradient of $\tilde{f}_t$ at x_t. The first term $\sum_{t=1}^T \langle \hat{g}_t, x_t - x^* \rangle$ of the right-hand side of (23) can be analyzed via Lemma 4.2, a part of which can be bounded with the following:

Lemma 4.9. *If ψ_t is defined by (19) with λ_t satisfying $|\frac{\hat{g}_{ti} v_{ti}}{\lambda_t}| \leq \frac{1}{2}$ for all $i \in [n]$, we have*

$$\langle \hat{g}_t, x_t - x_{t+1} \rangle - D_t(x_{t+1}, x_t) \leq \frac{1}{\lambda_t} \sum_{i=1}^{n} \hat{g}_{ti}^2 v_{ti}^2, \quad (24)$$

where v_{ti} is defined in (21). In addition, if $\hat{g}_t$ is given by (18), the expectation of the right-hand side of (24) is at most $O(\frac{n}{\gamma_t \lambda_t}\|v_t\|_2^2)$.

Combining this lemma with Lemmas 4.2 and 4.8, we obtain

$$R_T \leq O\left(\mathbf{E}\left[\sum_{t=1}^{T}\left(\frac{n\|v_t\|_2^2}{\gamma_t \lambda_t} + \gamma_t\right) + n\lambda_{T+1}\log T\right]\right).$$

From this and the definitions (22) of parameters γ_t and λ_t, we have the following:

Proposition 4.10. *The regret for Algorithm 2 is bounded as follows:*

$$R_T = O\left(\mathbf{E}\left[(n^2 \log T)^{1/3}\left(\sum_{t=1}^{T}\|v_t\|_2\right)^{2/3}\right] + n^2 \log T\right).$$

We are now ready to show Theorem 4.7.

Proof of Theorem 4.7. We can show the regret bound of $O(nT^{2/3}(\log T)^{1/3})$ from $\|v_t\|_2 = O(\sqrt{n})$, in a similar way to the proof of Theorem 4.3.

As we have $\|v_t\|_2 \leq \sqrt{n}\|v_t\|_\infty$, we have $\sum_{t=1}^T \|v_t\|_2 \leq \sqrt{n}\sum_{t=1}^T \|v_t\|_\infty := \sqrt{n}Q$. Hence, from Proposition 4.10, there exists a constant B such that $R_T \leq B(n(\log T)^{1/3}Q^{2/3} + n^2 \log T)$. Combining this with Lemma 4.6, for any $\eta \in (0, 1]$, we obtain

$$\begin{aligned} R_T &= (1+\eta)R_T - \eta R_T \\ &= (1+\eta)B(n(\log T)^{1/3}Q^{2/3} + n^2 \log T) - \eta(\Delta Q - 2C) \\ &\leq 2Bn(\log T)^{1/3}Q^{2/3} - \eta\Delta Q + 2n^2 \log T + 2\eta C. \end{aligned} \quad (25)$$

We then use the inequality of

$$ax^2 - bx^3 \leq \frac{4a^3}{27b^2}. \quad (26)$$

that holds for $a \geq 0, b > 0$ and $x > 0$. We can confirm (26) by considering the maximizer of the left-hand side. In fact, the left-hand side is maximized when $2ax - 3bx^2 = 0$, which means, $x = \frac{2a}{3b}$. Then the left-hand side is equal to $\frac{4a^3}{27b^3}$, which implies that (26) holds for any $x > 0$. Applying (26) with $a = 2Bn(\log T)^{1/3}$, $b = \eta\Delta$, and $x = Q^{1/3}$ to (25), we obtain $R_T \leq \frac{32n^3 \log T}{27\eta^2\Delta^2} + 2n^2 \log T + 2\eta C = O\left(\frac{n^3 \log T}{\eta^2\Delta^2} + \eta C\right)$. By setting $\eta = \min\left\{1, C^{-1/3}\left(\frac{n^3 \log T}{\Delta^2}\right)^{1/3}\right\}$, we obtain the regret bounds for the stochastic model (with adversarial corruptions) in Theorem 4.7. □

5. Lower Bounds

This section discusses regret lower bounds for online submodular minimization. For the worst-case regret bounds of $\Omega(\sqrt{nT})$ in the full-information setting and of $\Omega(n\sqrt{T})$ in the bandit-feedback setting, see, e.g., Theorem 14 by Hazan & Kale (2012) and Theorem 3 by Dani et al. (2008), respectively.

The gap- and corruption-level-dependent regret lower bounds of $\Omega(\frac{1}{\Delta} + \sqrt{\frac{C}{\Delta}})$ for the full-information setting immediately follow from Theorem 5 by Ito (2021b) as online submodular minimization includes the problem of prediction with experts (with $N = 2$ experts) as a special case. For the bandit-feedback setting, we can show a lower bound of $\Omega(\frac{n}{\Delta} + \sqrt{\frac{Cn}{\Delta}})$ by combining the techniques in the proof of Theorem 3 in the paper (Dani et al., 2008) and those for Theorem 5 in (Ito, 2021b).

Theorem 5.1. *For any $\Delta \in (0, \frac{1}{4n})$, $n \geq 4$, $T \geq n^2$ and $C \in [0, T]$, and for any online submodular minimization algorithm with bandit feedback, there exists an environment in the stochastic model with adversarial corruption with the given parameters Δ, n, T, and C, for which the regret is bounded from below as $R_T = \Omega(\min\{\frac{n}{\Delta} + \sqrt{\frac{Cn}{\Delta}}, n\sqrt{T}\})$.*

6. Conclusions And Open Questions

This paper revisits online submodular minimization and provides best-of-both-worlds algorithms with gap-dependent regret bounds and robustness to adversarial corruption. In both settings of full information and bandit feedback, there are gaps between the upper and lower bounds, and closing these gaps remains open problems. The lower bounds in this paper are constructed with the environment with linear objectives f_t. For such special cases of linear optimization problems, Zimmert & Seldin (2019) have provided

an algorithm with bandit feedback, of which regret upper bound matches the lower bounds, as shown in their Theorem 4. This fact means that we need to consider the problem instances with nonlinear objective functions in order to improve the lower bound. On the other hand, there is still the possibility that the lower bounds are tight, and thus online submodular minimization is only as difficult as online linear optimization.

Acknowledgements

The author was supported by JST, ACT-I Grant Number JPMJPR18U5, Japan.

References

Amir, I., Attias, I., Koren, T., Mansour, Y., and Livni, R. Prediction with corrupted expert advice. *Advances in Neural Information Processing Systems*, 33, 2020.

Auer, P. and Chiang, C.-K. An algorithm with nearly optimal pseudo-regret for both stochastic and adversarial bandits. In *Conference on Learning Theory*, pp. 116–120, 2016.

Axelrod, B., Liu, Y. P., and Sidford, A. Near-optimal approximate discrete and continuous submodular function minimization. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 837–853. SIAM, 2020.

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

Bilmes, J. Submodularity in machine learning and artificial intelligence. *arXiv preprint arXiv:2202.00132*, 2022.

Bogunovic, I., Krause, A., and Scarlett, J. Corruption-tolerant gaussian process bandit optimization. In *International Conference on Artificial Intelligence and Statistics*, pp. 1071–1081, 2020.

Bogunovic, I., Losalka, A., Krause, A., and Scarlett, J. Stochastic linear bandits robust to adversarial attacks. In *International Conference on Artificial Intelligence and Statistics*, pp. 991–999, 2021.

Bubeck, S. and Slivkins, A. The best of both worlds: Stochastic and adversarial bandits. In *Conference on Learning Theory*, pp. 42–1, 2012.

Cesa-Bianchi, N. and Lugosi, G. *Prediction, Learning, and Games*. Cambridge university press, 2006.

Chakrabarty, D., Lee, Y. T., Sidford, A., and Wong, S. C.-w. Subquadratic submodular function minimization. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pp. 1220–1231. ACM, 2017.

Chen, L., Harshaw, C., Hassani, H., and Karbasi, A. Projection-free online optimization with stochastic gradient: From convexity to submodularity. In *International Conference on Machine Learning*, pp. 814–823. PMLR, 2018a.

Chen, L., Hassani, H., and Karbasi, A. Online continuous submodular maximization. In *International Conference on Artificial Intelligence and Statistics*, pp. 1896–1905. PMLR, 2018b.

Dani, V., Hayes, T. P., and Kakade, S. M. Stochastic linear optimization under bandit feedback. In *Conference on Learning Theory*, pp. 355–366, 2008.

De Rooij, S., Van Erven, T., Grünwald, P. D., and Koolen, W. M. Follow the leader if you can, hedge if you must. *The Journal of Machine Learning Research*, 15(1):1281–1316, 2014.

Degenne, R. and Perchet, V. Anytime optimal algorithms in stochastic multi-armed bandits. In *International Conference on Machine Learning*, pp. 1587–1595. PMLR, 2016.

Erez, L. and Koren, T. Towards best-of-all-worlds online learning with feedback graphs. *Advances in Neural Information Processing Systems*, 34:28511–28521, 2021.

Fujishige, S. *Submodular Functions and Optimization*, volume 58. Elsevier, 2005.

Gaillard, P., Stoltz, G., and Van Erven, T. A second-order bound with excess losses. In *Conference on Learning Theory*, pp. 176–196. PMLR, 2014.

Grötschel, M., Lovász, L., and Schrijver, A. The ellipsoid method and its consequences in combinatorial optimization. *Combinatorica*, 1(2):169–197, 1981.

Gupta, A., Koren, T., and Talwar, K. Better algorithms for stochastic bandits with adversarial corruptions. In *Conference on Learning Theory*, pp. 1562–1578, 2019.

Hajiesmaili, M., Talebi, M. S., Lui, J., Wong, W. S., et al. Adversarial bandits with corruptions: Regret lower bound and no-regret algorithm. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

Harvey, N., Liaw, C., and Soma, T. Improved algorithms for online submodular maximization via first-order regret bounds. *Advances in Neural Information Processing Systems*, 33, 2020.

Hazan, E. and Kale, S. Online submodular minimization. *Journal of Machine Learning Research*, 13(Oct):2903–2922, 2012.

Ito, S. Submodular function minimization with noisy evaluation oracle. *Advances in Neural Information Processing Systems*, 32:12103–12113, 2019.

Ito, S. Hybrid regret bounds for combinatorial semi-bandits and adversarial linear bandits. *Advances in Neural Information Processing Systems*, 34, 2021a.

Ito, S. On optimal robustness to adversarial corruption in online decision problems. *Advances in Neural Information Processing Systems*, 34, 2021b.

Ito, S. Parameter-free multi-armed bandit algorithms with hybrid data-dependent regret bounds. In *Conference on Learning Theory*, pp. 2552–2583. PMLR, 2021c.

Ito, S. and Fujimaki, R. Large-scale price optimization via network flow. In *Advances in Neural Information Processing Systems*, pp. 3855–3863, 2016.

Iwata, S. A faster scaling algorithm for minimizing submodular functions. *SIAM Journal on Computing*, 32(4): 833–840, 2003.

Iwata, S. Submodular function minimization. *Mathematical Programming*, 112(1):45–64, 2008.

Iwata, S., Fleischer, L., and Fujishige, S. A combinatorial strongly polynomial algorithm for minimizing submodular functions. *Journal of the ACM (JACM)*, 48(4): 761–777, 2001.

Jegelka, S. and Bilmes, J. Online submodular minimization for combinatorial structures. In *International Conference on Machine Learning*, pp. 345–352, 2011a.

Jegelka, S. and Bilmes, J. Submodularity beyond submodular energies: Coupling edges in graph cuts. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1897–1904, 2011b.

Jin, T. and Luo, H. Simultaneously learning stochastic and adversarial episodic mdps with known transition. *Advances in Neural Information Processing Systems*, 33, 2020.

Jin, T., Huang, L., and Luo, H. The best of both worlds: stochastic and adversarial episodic mdps with unknown transition. *Advances in Neural Information Processing Systems*, 34, 2021.

Jun, K.-S., Li, L., Ma, Y., and Zhu, X. Adversarial attacks on stochastic bandits. In *Advances in Neural Information Processing Systems*, pp. 3644–3653, 2018.

Kohli, P. and Torr, P. H. Dynamic graph cuts and their applications in computer vision. In *Computer Vision*, pp. 51–108. Springer, 2010.

Lattimore, T. and Szepesvári, C. *Bandit Algorithms*. Cambridge University Press, 2020.

Lee, C.-W., Luo, H., Wei, C.-Y., Zhang, M., and Zhang, X. Achieving near instance-optimality and minimax-optimality in stochastic and adversarial linear bandits simultaneously. In *International Conference on Machine Learning*, 2021.

Lee, Y. T., Sidford, A., and Wong, S. C.-w. A faster cutting plane method and its implications for combinatorial and convex optimization. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pp. 1049–1065. IEEE, 2015.

Lin, H. and Bilmes, J. An application of the submodular principal partition to training data subset selection. In *NIPS workshop on Discrete Optimization in Machine Learning*. Citeseer, 2010.

Liu, F. and Shroff, N. Data poisoning attacks on stochastic bandits. In *International Conference on Machine Learning*, pp. 4042–4050, 2019.

Lovász, L. Submodular functions and convexity. In *Mathematical Programming The State of the Art*, pp. 235–257. Springer, 1983.

Luo, H. and Schapire, R. E. Achieving all with no parameters: Adanormalhedge. In *Conference on Learning Theory*, pp. 1286–1304. PMLR, 2015.

Lykouris, T., Mirrokni, V., and Paes Leme, R. Stochastic bandits robust to adversarial corruptions. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pp. 114–122, 2018.

Masoudian, S. and Seldin, Y. Improved analysis of the tsallis-inf algorithm in stochastically constrained adversarial bandits and stochastic bandits with adversarial corruptions. In *Conference on Learning Theory*, pp. 3330–3350. PMLR, 2021.

Matsuoka, T., Ito, S., and Ohsaka, N. Tracking regret bounds for online submodular optimization. In *International Conference on Artificial Intelligence and Statistics*, pp. 3421–3429. PMLR, 2021.

McCormick, S. T. Submodular function minimization. *Handbooks in operations research and management science*, 12:321–391, 2005.

Mourtada, J. and Gaïffas, S. On the optimality of the hedge algorithm in the stochastic regime. *Journal of Machine Learning Research*, 20:1–28, 2019.

Orlin, J. B. A faster strongly polynomial time algorithm for submodular function minimization. *Mathematical Programming*, 118(2):237–251, 2009.

Roughgarden, T. and Wang, J. R. An optimal learning algorithm for online unconstrained submodular maximization. In *Conference On Learning Theory*, pp. 1307–1325. PMLR, 2018.

Schrijver, A. A combinatorial algorithm minimizing submodular functions in strongly polynomial time. *Journal of Combinatorial Theory, Series B*, 80(2):346–355, 2000.

Seldin, Y. and Slivkins, A. One practical algorithm for both stochastic and adversarial bandits. In *International Conference on Machine Learning*, pp. 1287–1295, 2014.

Soma, T. No-regret algorithms for online k-submodular maximization. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pp. 1205–1214. PMLR, 2019.

Streeter, M. and Golovin, D. An online algorithm for maximizing submodular functions. In *Proceedings of the 21st International Conference on Neural Information Processing Systems*, pp. 1577–1584, 2008.

Wei, C.-Y. and Luo, H. More adaptive algorithms for adversarial bandits. In *Conference On Learning Theory*, pp. 1263–1291, 2018.

Zhang, M., Chen, L., Hassani, H., and Karbasi, A. Online continuous submodular maximization: from full-information to bandit feedback. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, pp. 9210–9221, 2019.

Zimmert, J. and Seldin, Y. An optimal algorithm for stochastic and adversarial bandits. In *International Conference on Artificial Intelligence and Statistics*, pp. 467–475, 2019.

Zimmert, J. and Seldin, Y. Tsallis-inf: An optimal algorithm for stochastic and adversarial bandits. *Journal of Machine Learning Research*, 22(28):1–49, 2021.

A. Omitted Proofs

A.1. Proof of Lemma 4.4

We first show

$$\langle g_t, x_t - x_{t+1} \rangle - D_t(x_{t+1}, x_t) \leq \sum_{i=1}^{n} \lambda_{ti} \cdot \nu \left(\frac{g_{ti}}{\lambda_{ti}}, x_{ti} \right). \tag{27}$$

When ψ_t is defined by (10), the Bregman divergence D_t associated with ψ_t can be expressed as

$$D_t(x_{t+1}, x_t) = \sum_{i=1}^{n} \lambda_{ti} d(x_{t+1,i}, x_{ti}),$$

where d is defined by

$$\begin{aligned} d(a,b) &= \phi(a) - \phi(b) - \phi'(b)(a-b) \\ &= a \log a + (1-a)\log(1-a) - b \log b - (1-b)\log(1-b) - (\log b - \log(1-b))(a-b) \\ &= a \log \frac{a}{b} + (1-a) \log \frac{1-a}{1-b}. \end{aligned}$$

Hence, we have

$$\langle g_t, x_t - x_{t+1} \rangle - D_t(x_{t+1}, x_t) = \sum_{i=1}^{n} \left(g_{ti} \cdot (x_{ti} - x_{t+1,i}) - \lambda_{ti} d(x_{t+1,i}, x_{ti}) \right). \tag{28}$$

The right-hand side of this can be bounded via the following:

$$\max_{y \in [0,1]} \{\xi(y)\} := \max_{y \in [0,1]} \{g \cdot (x-y) - \lambda d(y,x)\} = \lambda \cdot \nu \left(\frac{g}{\lambda}, x \right), \tag{29}$$

where ν is defined in (12). We can see that (29) holds for any $g \in \mathbb{R}$, $x \in (0,1)$, and $\lambda > 0$. In fact, ξ is a concave function and its derivative can be expressed as

$$\frac{\mathrm{d}\xi(y)}{\mathrm{d}y} = -g - \lambda \cdot (\phi'(y) - \phi'(x)) = -g - \lambda \cdot \left(\log \frac{y}{1-y} - \log \frac{x}{1-x} \right),$$

and hence, the maximum is attained when y satisfies $-g - \lambda \cdot (\phi'(y) - \phi'(x)) = 0$. When we substitute such y into $\xi(y)$, we have $\xi(y) = \lambda \cdot \nu(\frac{g}{\lambda}, x)$, which implies (29) holds. By combining (28) and (29), we obtain (27).

We next show

$$\nu(g,z) \leq \min\{z, 1-z\} g^2 \tag{30}$$

holds for any $g \in [-1/2, 1/2]$ and $z \in [0,1]$. As we have $\log(1+x) \leq x$ for any $x > -1$, we have

$$\nu(g,z) = \log(1 - z + z\exp(g)) - zg \leq -z + z\exp(g) - zg = z(\exp(g) - g - 1) \leq zg^2, \tag{31}$$

where the last inequality follows from $g \leq 1/2$. Similarly, we have

$$\begin{aligned} \nu(g,z) &= \log(1 - z + z\exp(g)) - zg = \log(\exp(-g) - z\exp(-g) + z) + (1-z)g \\ &= \log(1 + (1-z)(\exp(-g) - 1)) + (1-z)g \leq (1-z)(\exp(-g) - 1) + (1-z)g \\ &= (1-z)(\exp(-g) + g - 1) \leq (1-z)g^2, \end{aligned} \tag{32}$$

where the last inequality follows from $-g \leq 1/2$. From (31) and (32), we have (30). The inequality (27) combined with (30) completes the proof of Lemma 4.4. □

A.2. Proof of Proposition 4.5

From (4) and the definition of X_t in Algorithm 1, the regret can be expressed as

$$R_T = \mathbf{E}\left[\sum_{t=1}^{T} f_t(X_t) - \sum_{t=1}^{T} f_t(X^*)\right] = \mathbf{E}\left[\sum_{t=1}^{T} f_t(H_{u_t}(x_t)) - \sum_{t=1}^{T} \tilde{f}_t(x^*)\right] = \mathbf{E}\left[\sum_{t=1}^{T} \tilde{f}_t(x_t) - \sum_{t=1}^{T} \tilde{f}_t(x^*)\right],$$

where $X^* \in \arg\min_{X \in 2^{[n]}} \mathbf{E}[\sum_{t=1}^{T} f(X)]$ and $x^* \in \{0,1\}^n$ is the indicator vector of X^*, i.e., $x^*_i = \mathbf{1}[i \in X^*]$ for all $i \in [n]$. From Lemmas 4.2 and 4.4, $\sum_{t=1}^{T} \tilde{f}_t(x_t) - \sum_{t=1}^{T} \tilde{f}_t(x^*)$ can be bounded as

$$\begin{aligned}
\sum_{t=1}^{T} \left(\tilde{f}_t(x_t) - \tilde{f}_t(x^*)\right) &\le \sum_{t=1}^{T} \left(\langle g_t, x_t - x_{t+1}\rangle - D_t(x_{t+1}, x_t)\right) + \sum_{t=1}^{T} \left(\psi_t(x_{t+1}) - \psi_{t+1}(x_{t+1})\right) + \psi_{T+1}(x^*) - \psi_1(x_1) \\
&\le \sum_{t=1}^{T}\sum_{i=1}^{n} \lambda_{ti} \cdot \nu\left(\frac{g_{ti}}{\lambda_{ti}}, x_{ti}\right) + \sum_{t=1}^{T}\sum_{i=1}^{n} (\lambda_{ti} - \lambda_{t+1,i})\phi(x_{t+1,i}) - \psi_1(x_1) \\
&\le \sum_{t=1}^{T}\sum_{i=1}^{n} \lambda_{ti} \cdot \nu\left(\frac{g_{ti}}{\lambda_{ti}}, x_{ti}\right) + \log 2 \cdot \sum_{i=1}^{n} \lambda_{T+1,i} \le 2 \log 2 \cdot \sum_{i=1}^{n} \lambda_{T+1,i}, \qquad (33)
\end{aligned}$$

where the first inequality follows from Lemma 4.2, the second inequality follows from Lemma 4.4 and the definition (10) of ψ_t, the third inequality follows from $\psi(x) \in [-\log 2, 0]$ for any $x \in [0,1]$, and the last inequality follows from (11).

We next show that

$$\lambda_{ti} \le 2 + \sqrt{\frac{2}{\log 2} \sum_{s=1}^{t-1} g_{si}^2 \min\{x_{si}, 1 - x_{si}\}} \qquad (34)$$

holds for any i and t, by induction in t. If $t = 1$, (34) immediately follows from (11). From the definition (11) of λ_{ti} and from (30), we have

$$\lambda_{t+1,i} = \lambda_{t,i} + \frac{1}{\log 2} \lambda_{ti} \cdot \nu\left(\frac{g_{ti}}{\lambda_{ti}}, x_{ti}\right) \le \lambda_{ti} + \frac{g_{ti}^2 \{x_{ti}, 1 - x_{ti}\}}{\log 2 \cdot \lambda_{ti}}.$$

As the right-hand side of this is monotone non-decreasing in $\lambda_{ti} \ge 2$, if (34) holds for a given t, we have

$$\begin{aligned}
\lambda_{t+1,i} &\le 2 + \sqrt{\frac{2}{\log 2} \sum_{s=1}^{t-1} g_{si}^2 \min\{x_{si}, 1 - x_{si}\}} + \frac{g_{ti}^2 \{x_{ti}, 1 - x_{ti}\}}{\log 2 \cdot \left(2 + \sqrt{\frac{2}{\log 2} \sum_{s=1}^{t-1} g_{si}^2 \min\{x_{si}, 1 - x_{si}\}}\right)} \\
&\le 2 + \sqrt{\frac{2}{\log 2} \sum_{s=1}^{t-1} g_{si}^2 \min\{x_{si}, 1 - x_{si}\}} + \sqrt{\frac{2}{\log 2}} \cdot \frac{g_{ti}^2 \{x_{ti}, 1 - x_{ti}\}}{2\sqrt{\sum_{s=1}^{t} g_{si}^2 \min\{x_{si}, 1 - x_{si}\}}} \\
&\le 2 + \sqrt{\frac{2}{\log 2} \sum_{s=1}^{t-1} g_{si}^2 \min\{x_{si}, 1 - x_{si}\}} + \sqrt{\frac{2}{\log 2}} \cdot \left(\sqrt{\sum_{s=1}^{t} g_{si}^2 \min\{x_{si}, 1 - x_{si}\}} - \sqrt{\sum_{s=1}^{t-1} g_{si}^2 \min\{x_{si}, 1 - x_{si}\}}\right) \\
&\le 2 + \sqrt{\frac{2}{\log 2} \sum_{s=1}^{t} g_{si}^2 \min\{x_{si}, 1 - x_{si}\}},
\end{aligned}$$

which means that (34) holds even when we substitute $t + 1$ for t. Hence, by induction in $t \ge 1$, it was shown that (34) holds for all $t \ge 1$.

Combining (33) and (34), we obtain

$$\sum_{t=1}^{T}\left(\tilde{f}_t(x_t)-\tilde{f}_t(x^*)\right) \le 2\log 2\cdot\sum_{i=1}^{n}\lambda_{T+1,i} \le 4\log 2\cdot n+2\sqrt{2\log 2}\sum_{i=1}^{n}\sqrt{\sum_{t=1}^{T}g_{ti}^2\min\{x_{si},1-x_{ti}\}}$$
$$\le 4\log 2\cdot n+2\sqrt{2\log 2}\sqrt{n\sum_{t=1}^{T}\sum_{i=1}^{n}g_{ti}^2\min\{x_{si},1-x_{ti}\}} \le 4\log 2\cdot n+2\sqrt{2\log 2}\sqrt{n\sum_{t=1}^{T}\sum_{i=1}^{n}|g_{ti}|\min\{x_{si},1-x_{ti}\}}$$
$$\le 4\log 2\cdot n+2\sqrt{2\log 2}\sqrt{n\sum_{t=1}^{T}\left(\sum_{i=1}^{n}|g_{ti}|\right)\max_{i\in[n]}\{\min\{x_{si},1-x_{ti}\}\}}$$
$$\le 4\log 2\cdot n+4\sqrt{2\log 2}\sqrt{n\sum_{t=1}^{T}\max_{i\in[n]}\{\min\{x_{si},1-x_{ti}\}\}}, \tag{35}$$

where the third inequality follows from the Cauchy-Schwarz inequality, the forth inequality follows from $g_t\in[-1,1]^n$, and the last inequality follows from Lemma 4.1.

□

A.3. Proof of Lemma 4.6

From the definition of Δ in Section 3, if f'_t follows $\mathcal{D}$, we have

$$\mathbf{E}\left[\sum_{t=1}^{T}f'_t(X_t)-\sum_{t=1}^{T}f'_t(X^*)\right]=\mathbf{E}\left[\sum_{t=1}^{T}\left(\bar{f}(X_t)-\bar{f}(X^*)\right)\right]\ge\mathbf{E}\left[\Delta\sum_{t=1}^{T}\mathbf{1}[X_t\neq X^*]\right]=\Delta\sum_{t=1}^{T}\mathrm{Prob}[X_t\neq X^*]. \tag{36}$$

Hence, in the stochastic setting with adversarial corruptions, we have

$$\mathbf{E}\left[\sum_{t=1}^{T}f_t(X_t)-\sum_{t=1}^{T}f_t(X^*)\right]\le\mathbf{E}\left[\sum_{t=1}^{T}f'_t(X_t)-\sum_{t=1}^{T}f'_t(X^*)\right]+\sum_{t=1}^{T}\mathbf{E}\left[|f'_t(X_t)-f_t(X_t)|+|f'_t(X^*)-f_t(X^*)|\right]$$
$$\le\Delta\sum_{t=1}^{T}\mathrm{Prob}[X_t\neq X^*]+2C, \tag{37}$$

where the last inequality follows from (36), the definition (3) of the corruption level C, and the fact that f'_t and f_t are independent of X_t. Given $x_t\in[0,1]^n$, if X_t is given as $X_t=H_{u_t}(x_t)$ with $u_t\sim\mathrm{Unif}([0,1])$, we have

$$\mathrm{Prob}[X_t\neq X^*]=1-\mathrm{Prob}[X_t=X^*]=1-\max\left\{0,\min_{i\in X^*}x_{ti}-\max_{i\in[n]\setminus X^*}x_{ti}\right\}=\min\left\{1,\max_{i\in X^*}(1-x_{ti})+\max_{i\in[n]\setminus X^*}x_{ti}\right\}$$
$$\ge\max_{i\in X^*}(1-x_{ti})+\max_{i\in[n]\setminus X^*}x_{ti}\ge 2\max_{i\in[n]}\{\min\{x_{ti},1-x_{ti}\}\}.$$

By combining this with (37), we obtain the second inequality of Lemma 4.6. □

A.4. Proof of Lemma 4.8

From the definition of x^*, if u follows a uniform distribution over $[0,1]$, $H_u(x^*)=X^*$ with a probability at least $(1-\frac{2}{T})$. Hence, we have

$$\tilde{f}_t(x^*)\ge\left(1-\frac{2}{T}\right)f_t(X^*),$$

which implies

$$f_t(X^*)-\tilde{f}_t(x^*)\le\frac{2}{T}f_t(X^*)\le\frac{2}{T}.$$

Similarly, from the definition (16) of p_t, if X_t is given as in Algorithm 2, we have

$$\tilde{f}_t(x_t) - \mathbf{E}\left[f_t(X_t)\right] \leq \gamma_t$$

From the above two inequalities, we have

$$\begin{aligned} R_T &= \mathbf{E}\left[\sum_{t=1}^{T} f_t(X_t) - \sum_{t=1}^{T} f_t(X^*)\right] \leq \mathbf{E}\left[\sum_{t=1}^{T}(\tilde{f}_t(x_t) + \gamma_t) - \sum_{t=1}^{T}\left(\tilde{f}_t(x^*) - \frac{2}{T}\right)\right] \\ &= \mathbf{E}\left[\sum_{t=1}^{T}(\tilde{f}_t(x_t) - \tilde{f}_t(x^*)) - \sum_{t=1}^{T}\gamma_t\right] + 2. \end{aligned} \tag{38}$$

As $\tilde{f}_t$ is a convex function, for any subgradient g_t of $\tilde{f}_t$ at x_t, we have $\tilde{f}_t(x_t) - \tilde{f}_t(x^*) \leq \langle g_t, x_t - x^* \rangle$. Further, as $\hat{g}_t$ defined in (18) is an unbiased estimator of a subgradient of $\tilde{f}_t$, we have

$$\mathbf{E}\left[\tilde{f}_t(x_t) - \tilde{f}_t(x^*)\right] \leq \mathbf{E}\left[\langle \hat{g}_t, x_t - x^* \rangle\right].$$

By combining this with (38), we obtain the inequality of Lemma 4.8. □

A.5. Proof of Lemma 4.9

When ψ_t is defined by (19), the Bregman divergence D_t associated with ψ_t can be expressed as

$$D_t(x_{t+1}, x_t) = \lambda_t \sum_{i=1}^{n} d(x_{t+1,i}, x_{ti}),$$

where d is defined by

$$\begin{aligned} d(a,b) &= -\log a - \log(1-a) + \log b + \log(1-b) - \left(-\frac{1}{b} + \frac{1}{1-b}\right)(a-b) \\ &= -\log\frac{a}{b} + \frac{a-b}{b} - \log\frac{1-a}{1-b} + \frac{b-a}{1-b} = -\log\left(1 + \frac{a-b}{b}\right) + \frac{a-b}{b} - \log\left(1 + \frac{b-a}{1-b}\right) + \frac{b-a}{1-b} \\ &= \theta\left(\frac{a-b}{b}\right) + \theta\left(\frac{b-a}{1-b}\right). \end{aligned}$$

We here defined $\theta(x) = -\log(1+x) + x$. As $\log(1+x) \leq x$ holds for $x > -1$, we have $\theta(x) \geq 0$. From (39), we have

$$\begin{aligned} &\langle \hat{g}_t, x_t - x_{t+1} \rangle - D_t(x_{t+1}, x_t) = \sum_{i=1}^{n} \left(\hat{g}_{ti} \cdot (x_{ti} - x_{t+1,i}) - \lambda_t d(x_{t+1,i}, x_{ti})\right) \\ &= \sum_{i=1}^{n}\left(\hat{g}_{ti} \cdot (x_{ti} - x_{t+1,i}) - \lambda_t\left(\theta\left(\frac{x_{t+1,i} - x_{ti}}{x_{ti}}\right) + \theta\left(\frac{x_{ti} - x_{t+1,i}}{1 - x_{ti}}\right)\right)\right) \\ &\leq \sum_{i=1}^{n} \min\left\{\hat{g}_{ti} \cdot (x_{ti} - x_{t+1,i}) - \lambda_t\theta\left(\frac{x_{t+1,i} - x_{ti}}{x_{ti}}\right), \hat{g}_{ti} \cdot (x_{ti} - x_{t+1,i}) - \lambda_t\theta\left(\frac{x_{ti} - x_{t+1,i}}{1 - x_{ti}}\right)\right\}. \end{aligned} \tag{39}$$

The right-hand side of this can be bounded via the following:

$$\max_{y\in[0,1]}\{\xi_1(y)\} := \max_{y\in[0,1]}\left\{g \cdot (x-y) - \lambda\theta\left(\frac{y-x}{x}\right)\right\} = \lambda\theta\left(\frac{gx}{\lambda}\right) \leq \frac{(gx)^2}{\lambda}, \tag{40}$$

$$\max_{y\in[0,1]}\{\xi_2(y)\} := \max_{y\in[0,1]}\left\{g \cdot (x-y) - \lambda\theta\left(\frac{x-y}{1-x}\right)\right\} = \lambda\theta\left(\frac{g(1-x)}{\lambda}\right) \leq \frac{g(1-x)^2}{\lambda}, \tag{41}$$

We can see that (40) holds for any $g \in \mathbb{R}$, $x \in (0,1)$, and λ such that $|\frac{gx}{\lambda}| \leq \frac{1}{2}$. In fact, ξ_1 is a concave function and its derivative can be expressed as

$$\frac{\mathrm{d}\xi_1(y)}{\mathrm{d}y} = -g - \lambda \cdot \left(-\frac{1}{y} + \frac{1}{x}\right)$$

and hence, the maximum is attained when y satisfies $\frac{1}{y} = \frac{1}{x} + \frac{g}{\lambda}$. We then have

$$\begin{aligned}\xi_1(y) &= g \cdot (x-y) - \lambda\theta\left(\frac{y-x}{x}\right) = g \cdot (x-y) - \lambda \cdot \left(-\log\frac{y}{x} + \frac{y-x}{x}\right) \\ &= (x-y)\left(g + \frac{\lambda}{x}\right) - \lambda \log\frac{x}{y} = (x-y)\frac{\lambda}{y} - \lambda\log\frac{x}{y} = \lambda\left(-\log\frac{x}{y} + \frac{x}{y} - 1\right) \\ &= \lambda\left(-\log\left(1 + \frac{gx}{\lambda}\right) + \frac{gx}{\lambda}\right) = \lambda\theta\left(\frac{gx}{\lambda}\right) \le \frac{(gx)^2}{\lambda},\end{aligned}$$

where the last inequality follows from $|\frac{gx}{\lambda}| \le \frac{1}{2}$. This implies (40) holds. The other inequality of (41), which holds for any $g \in \mathbb{R}$, $x \in (0,1)$, and λ satisfying $|\frac{gx}{\lambda}| \le \frac{1}{2}$, can be shown in a similar way as well. By combining (39), (40) and (41), we obtain

$$\begin{aligned}\langle \hat{g}_t, x_t - x_{t+1}\rangle - D_t(x_{t+1}, x_t) &\le \sum_{i=1}^n \min\left\{\frac{(\hat{g}_{ti}x_{ti})^2}{\lambda_t}, \frac{(\hat{g}_{ti}(1-x_{ti}))^2}{\lambda_t}\right\} \\ &= \frac{1}{\lambda_t}\sum_{i=1}^n \hat{g}_{ti}^2 \min\left\{x_{ti}^2, (1-x_{ti})^2\right\} = \frac{1}{\lambda_t}\sum_{i=1}^n \hat{g}_{ti}^2 v_{ti}^2,\end{aligned}$$

which completes the proof of Lemma 4.9. □

A.6. Proof of Proposition 4.10

From Lemmas 4.2, 4.8 and 4.9, we have

$$\begin{aligned}R_T &\le \mathbf{E}\left[\sum_{t=1}^T \langle \hat{g}_t, x_t - x^*\rangle + 2\sum_{t=1}^T \gamma_t\right] + 4 \\ &\le \mathbf{E}\left[\sum_{t=1}^T (\langle \hat{g}_t, x_t - x_{t+1}\rangle - D_t(x_{t+1}, x_t)) + \sum_{t=1}^T (\psi_t(x_{t+1}) - \psi_{t+1}(x_{t+1})) + \psi_{T+1}(x^*) - \psi_1(x_1) + 2\sum_{t=1}^T \gamma_t\right] + 4 \\ &\le \mathbf{E}\left[\sum_{t=1}^T \frac{1}{\lambda_t}\sum_{i=1}^n \hat{g}_{ti}^2 v_{ti}^2 + \sum_{t=1}^T (\psi_t(x_{t+1}) - \psi_{t+1}(x_{t+1})) + \psi_{T+1}(x^*) + 2\sum_{t=1}^T \gamma_t\right] + 4, \qquad (42)\end{aligned}$$

where the first inequality follows from Lemma 4.8, the second inequality follows from Lemma 4.2, and the last inequality follows from Lemma 4.9, and the fact that ψ_t defined by (19) satisfies $\psi_t(x) \ge 0$ for any $x \in (0,1)^n$. From the definition of $\hat{g}_{ti}$ in (18), given x_t and p_t we have

$$\mathbf{E}\left[\hat{g}_{ti}^2\right] \le \mathbf{E}\left[\frac{1}{(p_t(i_t))^2}([\rho_{i_t}(\sigma_t)]_i)^2\right] = \sum_j^n p_t(j) \cdot \frac{1}{(p_t(j))^2}([\rho_j(\sigma_t)]_i)^2 \le 2\max_{j\in[n]}\frac{1}{p_t(j)} \le \frac{2(n+1)}{\gamma_t},$$

where the second inequality follows from the fact that $\rho_j(\sigma_t) \in \{-1, 0, 1\}^n$ has at most two non-zero entries, and the last inequality follows from (16). Hence, we have

$$\mathbf{E}\left[\sum_{i=1}^n \hat{g}_{ti}^2 v_{ti}^2\right] \le \mathbf{E}\left[\frac{2(n+1)}{\gamma_t}\sum_{i=1}^n v_{ti}^2\right] = \mathbf{E}\left[\frac{2(n+1)\|v_t\|_2^2}{\gamma_t}\right]. \qquad (43)$$

Further, we have

$$\begin{aligned}&\sum_{t=1}^T (\psi_t(x_{t+1}) - \psi_{t+1}(x_{t+1})) + \psi_{T+1}(x^*) \\ &= \sum_{t=1}^T (\lambda_t - \lambda_{t+1})\sum_{i=1}^n \left(\log\frac{1}{x_{t+1,i}} + \log\frac{1}{1-x_{t+1,i}}\right) + \lambda_{T+1}\sum_{i=1}^n \left(\log\frac{1}{x_i^*} + \log\frac{1}{1-x_i^*}\right) \\ &\le \lambda_{T+1}\sum_{i=1}^n \left(\log\frac{1}{x_i^*} + \log\frac{1}{1-x_i^*}\right) \le n\lambda_{T+1}\left(\log\frac{1}{1/T} + \log\frac{1}{1-1/T}\right) \le 2n\lambda_{T+1}\log T,\end{aligned}$$

where the first inequality follows from the fact that λ_t, defined in (22), is monotone non-decreasing, and the last second inequality follows from the definition of x^* in Lemma 4.8.

Combining this with (42) and (43), we obtain

$$R_T \leq 2\,\mathbf{E}\left[\sum_{t=1}^{T}\frac{(n+1)\|v_t\|_2^2}{\lambda_t\gamma_t} + n\lambda_{T+1}\log T + \sum_{t=1}^{T}\gamma_t\right] + 4 \leq 2\,\mathbf{E}\left[3\sum_{t=1}^{T}\frac{\sqrt{n}\|v_t\|_2}{\sqrt{\lambda_t}} + n\lambda_{T+1}\log T\right] + 4$$
$$\leq O\left(\mathbf{E}\left[(n^2\log T)^{1/3}\left(\sum_{t=1}^{T}\|v_t\|_2\right)^{2/3}\right] + n^2\log T\right),$$

where the second and the third inequalities follow from the definitions of γ_t and λ_t in (22), respectively. □

A.7. Proof of Theorem 5.1

We use the following lemma:

Lemma A.1 (Lemmas 3 and 4 in (Ito, 2019))**.** *For any $0 \leq \Delta \leq \frac{1}{4n}$ and $T \leq \frac{1}{8\Delta^2}$, and for any online submodular minimization algorithm with bandit feedback, there exists $X^* \subseteq [n]$ and a distribution $\mathcal{D}_{X^*,\Delta}$ of submodular functions such that $R_T = \Omega\left(\frac{n}{\Delta}\right)$ for $\{f_t\}_{t=1}^T \sim \mathcal{D}_\Delta^T$, and $\bar{f}(X) = \mathbf{E}_{f\sim\mathcal{D}_{X^*,\Delta}}[f(X)]$ satisfies $\bar{f}(X) = \frac{1}{2} + \frac{\Delta}{2}\left(2|(X^* \setminus X) \cup (X \setminus X^*)| - n\right)$ for some $X^* \subseteq [n]$.*

We note that the suboptimality gap defined in Section 3 for the distribution $\mathcal{D}_{X^*,\Delta}$ in this lemma is equal to Δ. Using this regret lower bound, we can show Theorem 5.1.

Proof of Theorem 5.1. We consider the following four cases, similarly to the proof of Theorem 5 by Ito (2021b): (i) If $T \leq \frac{1}{8\Delta^2}$, $\bar{R}_T = \Omega(n\sqrt{T})$. (ii) If $\frac{C}{n\Delta} \leq \frac{1}{8\Delta^2} \leq T$, $\bar{R}_T = \Omega(\frac{n}{\Delta})$. (iii) If $\frac{1}{8\Delta^2} \leq \frac{C}{n\Delta} \leq T$, $\bar{R}_T = \Omega(\sqrt{\frac{Cn}{\Delta}})$. (iv) If $\frac{1}{8\Delta^2} \leq T \leq \frac{C}{n\Delta}$, $\bar{R}_T = \Omega(n\sqrt{T})$. Combining all cases of (i)–(iv), we obtain the lower bound in Thoerem 5.1

(i) Suppose that $T \leq \frac{1}{8\Delta^2}$ holds. Set $\Delta' = \sqrt{\frac{1}{8T}}$. We then have $T = \frac{1}{8\Delta'^2}$ and $\Delta < \Delta' \leq \frac{1}{4n}$. Let $\mathcal{D} = \mathcal{D}_{X^*,\Delta'}$ be a distribution given in Lemma A.1 with $\Delta = \Delta'$. If $f_t \sim \mathcal{D}$ for all $t \in [T]$, then the environment is in the SwA model with the given parameters, and the regret is bounded as $R_T = \Omega(\frac{n}{Delta'}) = \Omega(n\sqrt{T})$ from Lemma 5.1.

(ii) Suppose that $\frac{C}{n\Delta} \leq \frac{1}{8\Delta^2} \leq T$ holds. If $f_t \sim \mathcal{D}_{X^*,\Delta}$ for all $t \in [T]$, the regret is bounded as $R_T = \Omega(\frac{n}{\Delta})$ from Lemma 5.1. From the condition of $\bar{f}$ given by $\mathcal{D}_{X^*,\Delta}$ in Lemma 5.1, the environment is in the SwA model with the given parameters.

(iii) Suppose that $\frac{1}{8\Delta^2} \leq \frac{C}{n\Delta} \leq T$ holds. Define $\Delta' = \sqrt{\frac{n\Delta}{8C}} \leq \Delta$. We then have $\frac{n}{\Delta'} = \sqrt{\frac{8nC}{\Delta}}$. Let $T' = \lceil\frac{1}{8\Delta'^2}\rceil = \lceil\frac{C}{n\Delta}\rceil \leq T$. Consider an environment in which $f_t \sim \mathcal{D}_{X^*,\Delta'}$ (distribution given in Lemma A.1 with $\Delta = \Delta'$) for $t \in [T']$, $f_t \sim \mathcal{D}_{X^*,\Delta}$ for $t \geq T'+1$. Then from Lemma A.1, we have $R_T \geq \bar{R}_{T'} = \Omega(\frac{n}{\Delta'}) = \Omega(\sqrt{\frac{Cn}{\Delta}})$. Further, we can show that the environment is in the SwA model with the given parameters. In fact, we have $nT'(\Delta - \Delta') \leq \frac{C}{\Delta}(\Delta - \Delta') \leq C$.

(iv) Suppose that $\frac{1}{8\Delta^2} \leq T \leq \frac{C}{n\Delta}$ holds, Set $\Delta' = \sqrt{\frac{1}{8T}} \leq \Delta$ and consider $f_t \sim \mathcal{D}_{X^*,\Delta'}$ for all $t \in [T]$. Then the regret is bounded as $R_T \geq \Omega(\frac{n}{\Delta'}) = \Omega(n\sqrt{T})$ from Lemma A.1. We can confirm that the environment is in the SwA model with the given parameters, as we have $n\Delta' T \leq n\Delta T \leq C$ from the assumptions on parameters. □